

STIC-Biotech/ChemLib

167494

119

From: Yu, Misook
Sent: Friday, September 30, 2005 4:58 PM
To: STIC-Biotech/ChemLib
Subject: 09/788,476

RECEIVED
OCT - 3 2005
STIC-BIOTECH DIVISION
(STIC)

Pls search SEQ ID NOs 1-3 including pending databases. Thank you,

Examiner Misook Yu, Ph.D.
571-272-0839 (Phone)
Art Unit 1642
REM-3A18 (Room)
REM-3C18 (Mail Box)

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 22:07:00 ; Search time 747.783 Seconds
(without alignments)
8254.755 Million cell updates/sec

Title: US-09-788-476A-1
Perfect score: 894
Sequence: 1 gggagtgagtgagggttaa.....taataaaaaaaatagaaaa 894

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	100.0	894	9	US-09-788-476A-1
2	894	100.0	1154	9	US-09-764-846-90
3	894	100.0	1154	14	US-10-091-483-90
4	873	97.7	873	9	US-09-788-476A-3
5	869.4	97.2	891	9	US-09-822-830A-389
6	833.8	93.3	1400	21	US-10-956-157-7827
7	833.8	93.3	1818	21	US-10-956-157-2592

8	733.4	82.0	3147	16	US-10-006-285-419	Sequence 419, App
9	690.8	77.3	2553	10	US-09-814-353-20673	Sequence 20673, A
c 10	595.2	66.6	2553	10	US-09-814-353-20673	Sequence 20673, A
11	498.4	55.7	1022	16	US-10-006-285-233	Sequence 233, App
12	460.8	51.5	488	17	US-10-242-535A-29658	Sequence 29658, A
13	460.8	51.5	488	18	US-10-085-783A-29658	Sequence 29658, A
14	458	51.2	470	17	US-10-242-535A-28856	Sequence 28856, A
15	458	51.2	470	18	US-10-085-783A-28856	Sequence 28856, A
16	431	48.2	470	17	US-10-242-535A-40257	Sequence 40257, A
17	431	48.2	470	18	US-10-085-783A-40257	Sequence 40257, A
18	389	43.5	426	17	US-10-242-535A-40571	Sequence 40571, A
19	389	43.5	426	18	US-10-085-783A-40571	Sequence 40571, A
20	379	42.4	379	17	US-10-242-535A-43248	Sequence 43248, A
21	379	42.4	379	18	US-10-085-783A-43248	Sequence 43248, A
22	336.2	37.6	463	10	US-09-918-995-15429	Sequence 15429, A
23	334	37.4	408	10	US-09-930-213-64	Sequence 64, Appl
c 24	281.4	31.5	5469	9	US-09-764-877-4000	Sequence 4000, Appl
c 25	281.4	31.5	5469	10	US-09-764-891-9371	Sequence 9371, Ap
c 26	281.4	31.5	5469	15	US-10-205-428-817	Sequence 817, App
c 27	281.4	31.5	5469	17	US-10-242-515-4000	Sequence 4000, Ap
c 28	281.4	31.5	9453	9	US-09-764-877-3999	Sequence 3999, Ap
c 29	281.4	31.5	9453	10	US-09-764-891-9370	Sequence 9370, Ap
c 30	281.4	31.5	9453	15	US-10-205-428-816	Sequence 816, App
c 31	281.4	31.5	9453	17	US-10-242-515-3999	Sequence 3999, Ap
c 32	246.2	27.5	300	16	US-10-006-285-196	Sequence 196, App
33	244	27.3	275	20	US-10-425-115-158622	Sequence 158622, A
34	211.8	23.7	255	10	US-09-930-213-531	Sequence 531, App
35	206	23.0	343	17	US-10-062-674-737	Sequence 737, App
36	174	19.5	257	17	US-10-062-674-1714	Sequence 1714, Ap
37	154.2	17.2	215	9	US-09-728-445-378	Sequence 378, App
38	154.2	17.2	215	22	US-10-964-549-378	Sequence 378, App
39	92.2	10.3	463	10	US-09-814-353-15759	Sequence 15759, A
40	55	6.2	399	22	US-10-450-763-26486	Sequence 26486, A
c 41	55	6.2	399	22	US-10-450-763-29210	Sequence 29210, A
42	54	6.0	5659	15	US-10-172-086-32	Sequence 32, Appl
43	54	6.0	5659	15	US-10-311-455-348	Sequence 348, App
44	54	6.0	5659	15	US-10-240-485-40	Sequence 40, Appl
45	54	6.0	5659	19	US-10-311-507-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-09-788-476A-1
; Sequence 1, Application US/09788476A
; Patent No. US20020107190A1
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Ching M. et al.
; TITLE OF INVENTION: NOVEL GENES AND EXPRESSION PRODUCTS THEREFROM
; FILE REFERENCE: 3669-0103P
; CURRENT APPLICATION NUMBER: US/09/788,476A
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 894
; TYPE: DNA
; ORGANISM: human
US-09-788-476A-1

Query Match 100.0%; Score 894; DB 9; Length 894;
Best Local Similarity 100.0%; Pred. No. 1.8e-229;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GGGAGTGGAGTGGGGTAAACAGATCGCACCGAGCGGTGGAGCTCCATAAGCTAAAG	60
Db	1	GGGAGTGGAGTGGGGTAAACAGATCGCACCGAGCGGTGGAGCTCCATAAGCTAAAG	60
Qy	61	CTTCCCGAACTAAAGCAAGATGCTTCTGCTGTTTGGAGACCAAGGGGATAAAGCAA	120
Db	61	CTTCCCGAACTAAAGCAAGATGCTTCTGCTGTTTGGAGACCAAGGGGATAAAGCAA	120

Qy	121	GATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATAGCTGAAGAGGAGCGCAATAGAA	180
Db	121	GATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATAGCTGAAGAGGAGCGCAATAGAA	180
Qy	181	GAAGATGTACTGGGAGATGAAACAGAGAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTTC	240
Db	181	GAAGATGTACTGGGAGATGAAACAGAGAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTTC	240
Qy	241	AAAGAGGAAAGAAACCCCTTGAAAAAACTCTTGATGTGGCAGCAGAGAAAGAAAGTGGTGAA	300
Db	241	AAAGAGGAAAGAAACCCCTTGAAAAAACTCTTGATGTGGCAGCAGAGAAAGAAAGTGGTGAA	300
Qy	301	ATTACATCTGAAATACACAGACTGAGAGAAATGCAGAGAGGGCTGAAACGATTCATATGA	360
Db	301	ATTACATCTGAAATACACAGACTGAGAGAAATGCAGAGAGGGCTGAAACGATTCATATGA	360
Qy	361	CCTGTGAGCTTCGGAGAGTAAGAAAGCTCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTT	420
Db	361	CCTGTGAGCTTCGGAGAGTAAGAAAGCTCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTT	420
Qy	421	CCAAACAAAAGGTCGTGCATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAAGGAA	480
Db	421	CCAAACAAAAGGTCGTGCATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAAGGAA	480
Qy	481	AGAGCTCAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG	540
Db	481	AGAGCTCAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG	540
Qy	541	AAACTGAAAAAGGAGGAGGCGATTTGGGATTTGTCAACAGTTCAGCTGCGAACTGGAACC	600
Db	541	AAACTGAAAAAGGAGGAGGCGATTTGGGATTTGTCAACAGTTCAGCTGCGAACTGGAACC	600
Qy	601	ACAGAGGATACAGAGGCAAGAGAGGAAAAAGAGCAGAGCGCTTTGGGATTTGCTCGATGA	660
Db	601	ACAGAGGATACAGAGGCAAGAGAGGAAAAAGAGCAGAGCGCTTTGGGATTTGCTCGATGA	660
Qy	661	AAAGTTCCCTGATACCTTCTGTTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTTGGTTCAC	720
Db	661	AAAGTTCCCTGATACCTTCTGTTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTTGGTTCAC	720
Qy	721	ATATATGCCTAAATGCCAGTCAATGTGCCTAGCTCTGCTCGCAGTACAGGAGGAGCATGTA	780
Db	721	ATATATGCCTAAATGCCAGTCAATGTGCCTAGCTCTGCTCGCAGTACAGGAGGAGCATGTA	780
Qy	781	CCCCAGGTACCTCACTGAACTCGGCGCAGAGTTTGACTTATTGCTGTTTTCAGCTTTTAAGG	840
Db	781	CCCCAGGTACCTCACTGAACTCGGCGCAGAGTTTGACTTATTGCTGTTTTCAGCTTTTAAGG	840
Qy	841	TTGTGCTGTTTTGTTTTGATTAATGTTGCTTGTGTTTAAATAAAAAAATAGAAA	894
Db	841	TTGTGCTGTTTTGTTTTGATTAATGTTGCTTGTGTTTAAATAAAAAAATAGAAA	894

RESULT 2			
US-09-764-846-90			
; Sequence 90, Application US/09764846			
; Patent No. US20020102638A1			
; GENERAL INFORMATION:			
; APPLICANT: Rosen et al.			
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies			
; FILE REFERENCE: PT212			
; CURRENT APPLICATION NUMBER: US/09/764,846			
; CURRENT FILING DATE: 2001-01-17			
; Prior application data removed - consult PALM or file wrapper			
; NUMBER OF SEQ ID NOS: 348			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 90			
; LENGTH: 1154			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: SITE			
; LOCATION: (1149)			
Db	837	AAAGTTCCTGATACATTTCTGTTCTCCAGTGTTCCTTCCTCTCTCTGGTCAC	896
Qy	721	ATATATGCCCTAAATGCCACAGTCATGTGCTACGTCCTCGCCTCGCAATGAGGAGCATGTGA	780
Db	897	ATATATGCCCTAAATGCCACAGTCATGTGCTACGTCCTCGCCTCGCAATGAGGAGCATGTGA	956
Qy	781	CCCCAGGTACATCCCATGAACTCGCGCAGCAGCTTTGACTTATTGCTGTTTTCAGCTTTAAAGG	840
Db	957	CCCCAGGTACATCCCATGAACTCGCGCAGCAGCTTTGACTTATTGCTGTTTTCAGCTTTAAAGG	1016
Qy	841	TTGTTGCTGTTTTGTTTTGTTTATGTTGCTTAAATAAAAAAATAGAAAA	894
Db	1017	TTGTTGCTGTTTTGTTTTGTTTATGTTGCTTAAATAAAAAAATAGAAAA	1070
RESULT 3			
US-10-091-483-90			
; Sequence 90, Application US/10091483			
; Publication No. US20030049650A1			
; GENERAL INFORMATION:			

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ12C1
; CURRENT APPLICATION NUMBER: US/10/091,483
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 348
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
; LENGTH: 1154
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1149)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-091-483-90

Query Match 100.0%; Score 894; DB 14; Length 1154;
Best Local Similarity 100.0%; Pred. No. 2,1e-229;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GGGAGTGGAGTGGGGGTAACAAGATGCGACCGAGCGGTGGAGCTCCATAAGCTAAAG 60
Db 177 GGGAGTGGAGTGGGGGTAACAAGATGCGACCGAGCGGTGGAGCTCCATAAGCTAAAG 236
Qy 61 CTTGCCGAACTAAAGCAAGAAATGCTTCTCTCGTGGTTGGAGACCAAGGGGAATAAGCAA 120
Db 237 CTTGCCGAACTAAAGCAAGAAATGCTTCTCTCGTGGTTGGAGACCAAGGGGAATAAGCAA 296
Qy 121 GATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAATGA 180
Db 297 GATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAATGA 356
Qy 181 GAAGATGTACTGGAGATGAACAGAGAGCAAGAAACAAAGCCCATTTGAGCTCCCTGTC 240
Db 357 GAAGATGTACTGGAGATGAACAGAGAGCAAGAAACAAAGCCCATTTGAGCTCCCTGTC 416
Qy 241 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTGA 300
Db 417 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTGA 476
Qy 301 ATTACATCTGAATATACCAAGACTGAGAGAAATGCAAGAGGGGCTGAAACGATCAATGTA 360
Db 477 ATTACATCTGAATATACCAAGACTGAGAGAAATGCAAGAGGGGCTGAAACGATCAATGTA 536
Qy 361 CTTGTGAGCTGGAGAGTAAGAAAGCTCTCGGGCAGCTAGCTTTGGGATTTCTTCAGTT 420
Db 537 CTTGTGAGCTGGAGAGTAAGAAAGCTCTCGGGCAGCTAGCTTTGGGATTTCTTCAGTT 596
Qy 421 CCACAAAGGCTCTGTCATCTGATAACAACTATGTTAACTTGGATAAGCTGAAGGA 480
Db 597 CCACAAAGGCTCTGTCATCTGATAACAACTATGTTAACTTGGATAAGCTGAAGGA 656
Qy 481 AGAGCTCAAGAAATTTGGTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 540
Db 657 AGAGCTCAAGAAATTTGGTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 716
Qy 541 AAATGAAAGAGAGGAAGAGGATTTGGGATTTGTCACAGTTGAGCTGGAACTGGAACC 600
Db 717 AAATGAAAGAGAGGAAGAGGATTTGGGATTTGTCACAGTTGAGCTGGAACTGGAACC 776
Qy 601 ACAGAGGATACAGAGGCAAGAGAGGAAGAGAGAGCGCTTTGGGATTTGCCTGATGA 660
Db 777 ACAGAGGATACAGAGGCAAGAGAGGAAGAGAGAGCGCTTTGGGATTTGCCTGATGA 836
Qy 661 AAAGTTCTGATACATTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTCTGTCAC 720
Db 837 AAAGTTCTGATACATTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTCTGTCAC 896
Qy 721 ATATATGCTTAAATGCACAGTCAATGTCCTACGTCCTGCTCGCAATGAGGAGCATGTA 780
Db 897 ATATATGCTTAAATGCACAGTCAATGTCCTACGTCCTGCTCGCAATGAGGAGCATGTA 956

Qy 781 CCCAGGTACATCCATGAACCTGCGCAGCGAGTTTGACTTATTCCTGTTTCAGCTTTAAGG 840
Db 957 CCCAGGTACATCCATGAACCTGCGCAGCGAGTTTGACTTATTCCTGTTTCAGCTTTAAGG 1016
Qy 841 TTGTTGTTGTTTTGTTTTGATTATGTTGCTGCTTAATAAAAAAATAGAAAA 894
Db 1017 TTGTTGTTGTTTTGTTTTGATTATGTTGCTGCTTAATAAAAAAATAGAAAA 1070
RESULT 4
US-09-788-476A-3
; Sequence 3, Application US/09788476A
; Patent No. US20020107190A1
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Ching M. et al.
; TITLE OF INVENTION: NOVEL GENES AND EXPRESSION PRODUCTS THEREFROM
; FILE REFERENCE: 3669-0103P
; CURRENT APPLICATION NUMBER: US/09/788,476A
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PCR Product
US-09-788-476A-3

Query Match 97.7%; Score 873; DB 9; Length 873;
Best Local Similarity 100.0%; Pred. No. 7.9e-224;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 6 TGGAGTGGAGGTAAACAAGATGCGACCGAGCGGTGGAGCTCCATAAGCTTAAAGCTTGC 65
Db 1 TGGAGTGGAGGTAAACAAGATGCGACCGAGCGGTGGAGCTCCATAAGCTTAAAGCTTGC 60
Qy 66 CGAACTAAAGCAAGATGTTCTGCTGCTGTTGGAGCAAGGAATAAAGCAAGATCT 125
Db 61 CGAACTAAAGCAAGATGTTCTGCTGCTGTTGGAGCAAGGAATAAAGCAAGATCT 120
Qy 126 TATCCACAGACTCAGGCATATCTTGAAGACATCTGTAAGAGGAGCGCAATGAAGAGA 185
Db 121 TATCCACAGACTCAGGCATATCTTGAAGACATCTGTAAGAGGAGCGCAATGAAGAGA 180
Qy 186 TGTACTGGAGATGAAACAGAGAGAAAGAAACAAAGCCCATTCAGCTCCCTGTCAAAGA 245
Db 181 TGTACTGGAGATGAAACAGAGAGAAAGAAACAAAGCCCATTCAGCTCCCTGTCAAAGA 240
Qy 246 GGAAGAAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAAGATTC 305
Db 241 GGAAGAAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAAGATTC 300
Qy 306 ATCTGAAATACCAAGACTGAGAGAAATGCAAGAGAGGGCTGAAACGATTCATGTACCTGT 365
Db 301 ATCTGAAATACCAAGACTGAGAGAAATGCAAGAGAGGGCTGAAACGATTCATGTACCTGT 360
Qy 366 GAGCTTGGAGAGTAAGAAAGCTCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAC 425
Db 361 GAGCTTGGAGAGTAAGAAAGCTCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAC 420
Qy 426 AAAAGGTCGTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAAGAGC 485
Db 421 AAAAGGTCGTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAAGAGC 480
Qy 486 TCAAAGATTTGGTTGAAATGCTCTTCAATCTCAGAAAGTCTGAAGATGATGAGAACT 545
Db 481 TCAAAGATTTGGTTGAAATGCTCTTCAATCTCAGAAAGTCTGAAGATGATGAGAACT 540
Qy 546 GAAAAGAGGAGAGCGATTTGGGATTTGTCACAAAGTTTCAGCTGGAACCTGCAACACAGA 605
Db 541 GAAAAGAGGAGAGCGATTTGGGATTTGTCACAAAGTTTCAGCTGGAACCTGCAACACAGA 600


```
Qy 178 CAAGAAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGAGCCCATTTGAGCTCCCT 237
Db 673 GAAGAAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGAGCCCATTTGAGCTCCCT 732
Qy 238 GTCAAGAGGAAGAAACCCCTTGAAACAACTGTCATGTGGCAGCAGAGGAAGAAAGTGGTG 297
Db 733 GTCAAGAGGAAGAAACCCCTTGAAACAACTGTCATGTGGCAGCAGAGGAAGAAAGTGGTG 792
Qy 298 AAAATTACATCTGAAATACCAACAGACTGAGAGAAATGCAAGAGGGCTGAAACGATTCAAT 357
Db 793 AAAATTACATCTGAAATACCAACAGACTGAGAGAAATGCAAGAGGGCTGAAACGATTCAAT 852
Qy 358 GTACTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCA 417
Db 853 GTACTGTGAGCTTGGAGAGTAAGAAAGTGTCTCGGCGAGCTAGGTTTGGGATTTCTTCA 912
Qy 418 GTTCCAAACAAAGGTCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAG 477
Db 913 GTTCCAAACAAAGGTCTGTCTCATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAG 972
Qy 478 GAAAGAGCTCAAGATTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAAGATGAT 537
Db 973 GAAAGAGCTCAAGATTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAAGATGAT 1032
Qy 538 GAGAAATCTGAAAGAGAGGAGGATTTGGGATTTGTCAAGATTGAGCTGGAACTGGA 597
Db 1033 GAGAAATCTGAAAGAGAGGAGGAGGATTTGGGATTTGTCAAGATTGAGCTGGAACTGGA 1092
Qy 598 ACCACAGAGGATACAGAGGCAAGAAAGAGGAAAGAGCAGCGCTTTGGGATTTGCCCTGA 657
Db 1093 ACCACAGAGGATACAGAGGCAAGAAAGAGGAAAGAGCAGCGCTTTGGGATTTGCCCTGA 1152
Qy 658 TGAAGAGTTCCTGATACCTTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTTCTTGGT 717
Db 1153 TGAAGAGTTCCTGATACCTTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTTCTTGGT 1212
Qy 718 CACATATATGCTTAATGACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGCAT 777
Db 1213 CACATATATGCTTAATGACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGCAT 1272
Qy 778 GTACCCAGGTACATCATGAACGAGCGAGCAGTGTGACTTATGCTGTTTCAAGCTTTA 837
Db 1273 GTACCCAGGTACATCATGAACGAGCGAGCAGTGTGACTTATGCTGTTTCAAGCTTTA 1332
Qy 838 AGGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 894
Db 1333 AGGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 1389
```

RESULT 7

```
US-10-956-157-2592
; Sequence 2592, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956.157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2592
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-2592
```

```
Query Match 93.3%; Score 833.8; DB 21; Length 1818;
Best Local Similarity 99.8%; Pred. No. 3.9e-213;
Matches 835; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 58 AAGTTCGCCGAACATAAGCAAGAATGTCTTGTCTGCTGCTTTGGAGACCAAGGGAATAAAG 117
Db 971 AAACTTGCCTGCAACATAAGCAAGAATGTCTTGTCTGCTGCTTTGGAGACCAAGGGAATAAAG 1030
Qy 118 CAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAAT 177
Db 1031 CAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAAT 1090
Qy 178 GAAGAAGATGTACTGGGAGATGAAACACAGAGAAAGAGAAACAAAGCCCATTTGAGCTCCCT 237
Db 1091 GAAGAAGATGTACTGGGAGATGAAACACAGAGAAAGAGAAACAAAGCCCATTTGAGCTCCCT 1150
Qy 238 GTCAAGAGGAAGAAACCCCTGAAACAACTGTTGATGTGCGACAGAGAAAGTGGTG 297
Db 1151 GTCAAGAGGAAGAAACCCCTGAAACAACTGTTGATGTGCGACAGAGAAAGTGGTG 1210
Qy 298 AAAATTACATCTGAAATACCAACAGACTGAGAGAAATGCAAGAGGGCTGAAACGATTCAAT 357
Db 1211 AAAATTACATCTGAAATACCAACAGACTGAGAGAAATGCAAGAGGGCTGAAACGATTCAAT 1270
Qy 358 GTACTGTGAGCTTGGAGAGTAAGAAAGCTGTCTGGGACAGTAGGTTTGGGATTTCTTCA 417
Db 1271 GTACTGTGAGCTTGGAGAGTAAGAAAGTGTCTGGGACAGTAGGTTTGGGATTTCTTCA 1330
Qy 418 GTTCCAAACAAAGGTCTGTCTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAG 477
Db 1331 GTTCCAAACAAAGGTCTGTCTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAG 1390
Qy 478 GAAAGAGCTCAAGATTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAAGATGAT 537
Db 1391 GAAAGAGCTCAAGATTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAAGATGAT 1450
Qy 538 GAGAAATCTGAAAGAGGAGGAGCGCATTTGGGATTTGTCAACAGTTCAGCTGGAACTGGA 597
Db 1451 GAGAAATCTGAAAGAGGAGGAGCGCATTTGGGATTTGTCAACAGTTCAGCTGGAACTGGA 1510
Qy 598 ACCACAGAGGATACAGAGGCAAGAAAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCCCTGA 657
Db 1511 ACCACAGAGGATACAGAGGCAAGAAAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCCCTGA 1570
Qy 658 TGAAGAGTTCCTGATACCTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTCTTGGT 717
Db 1571 TGAAGAGTTCCTGATACCTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTCTTGGT 1630
Qy 718 CACATATATGCTTAATGACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGCAT 777
Db 1631 CACATATATGCTTAATGACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGCAT 1690
Qy 778 GTACCCAGGTACATCATGAACCTGGGAGAGGAGTGTGACTTATGCTGTTTCAAGCTTTA 837
Db 1691 GTACCCAGGTACATCATGAACCTGGGAGAGGAGTGTGACTTATGCTGTTTCAAGCTTTA 1750
Qy 838 AGGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 894
Db 1751 AGGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 1807
```

RESULT 8

```
US-10-006-285-419
; Sequence 419, Application US/10006285
; Publication No. US20030165854A1
; GENERAL INFORMATION:
; APPLICANT: Mary Jane Cunningham
; APPLICANT: Matthew R. Kaser
; TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS
; FILE REFERENCE: PA-0039 US
; CURRENT APPLICATION NUMBER: US/10/006.285
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 514
; SOFTWARE: PERL Program
; SEQ ID NO 419
; LENGTH: 3147
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030165854A1 235169.27
US-10-006-285-419

Query Match      82.08; Score 733.4; DB 16; Length 3147;
Best Local Similarity 99.9%; Pred. No. 4.9e-186;
Matches 734; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 57 AAAGCTTGCCCAACTAAAGCAAGAAATCTTGTCTGCTGCTGTTGGAGACCAAGGGAATAA 116
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 97 AAAGCTTGCCCAACTAAAGCAAGAAATCTTGTCTGCTGCTGTTGGAGACCAAGGGAATAA 156
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 117 GCAAGATCTTATCACAAGCTCCAGGCATATCTTTGAAGAAACATGCTGAAGAGAGGCAAA 176
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 157 GCAAGATCTTATCACAAGCTCCAGGCATATCTTTGAAGAAACATGCTGAAGAGAGGCAAA 216
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 177 TGAGAGAGATGTAAGGAGATGAACACAGAGGAGAGAGAAACAAAGCCCATTCAGCTCCC 236
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 217 TGAAGAGAGATGTAAGGAGATGAACACAGAGGAGAGAGAAACAAAGCCCATTCAGCTCCC 276
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 237 TGTCAAAGAGAGAAACCCCTCGAAAAAATCTTTGATGTGCGACGACAGAGAAAGTGGT 296
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 277 TGTCAAAGAGAGAAACCCCTCGAAAAAATCTTTGATGTGCGACGACAGAGAAAGTGGT 336
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 297 GAAATTTACATCTGAAATACACAGACTGAGAGAAATCGAAGAGGGCTGAACGATTCAA 356
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 337 GAAATTTACATCTGAAATACACAGACTGAGAGAAATCGAAGAGGGCTGAACGATTCAA 396
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 357 TGACCTGTGAGCTTGAGAGATGAAGAGCTGCTCGGGCAGCTAGGTTGGGATTTCTTC 416
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 397 TGACCTGTGAGCTTGAGAGATGAAGAGCTGCTCGGGCAGCTAGGTTGGGATTTCTTC 456
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 417 AGTTCCAAACAAAGCTGTCTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAA 476
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 457 AGTTCCAAACAAAGCTGTCTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAA 516
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 477 GGAAGAGCTCAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGA 536
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 517 GGAAGAGCTCAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGA 576
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 537 TGAGAAACTGAAAGAGGAGGAGCGATTTGGGATTTGTACAGTTTCACTGGAAGTGG 596
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 577 TGAGAAACTGAAAGAGGAGGAGCGATTTGGGATTTGTACAGTTTCACTGGAAGTGG 636
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 597 AACCACAGAGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 656
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 637 AACCACAGAGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 696
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 657 ATGAAAAGTTCCTGATATTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTTGG 716
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 697 ATGAAAAGTTCCTGATATTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTTCTTTGG 756
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 717 TCACATATATCCCTAAATGCAAGTCTATGTCCTACCTGCTGCTGCTGCTGCTGCTGCTGCT 776
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 757 TCACATATATCCCTAAATGCAAGTCTATGTCCTACCTGCTGCTGCTGCTGCTGCTGCTGCT 816
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 777 TGTACCCCAAGGTACA 791
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 817 TGTACCCCAAGGTACA 831
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-814-353-20673
; Sequence 20673, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
```

```
Qy 758 GCCTCGCAATGAGGAGCATGTACCCAGGTACA 791
Db 1375 GCCTCGCAATGAGGAGCATGTACCCAGGTCA 1408

RESULT 10
US-09-814-353-20673/c
; Sequence 20673, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20673
; LENGTH: 2553
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-20673

Query Match 66.6%; Score 595.2; DB 10; Length 2553;
Best Local Similarity 96.9%; Pred. No. 6.1e-149;
Matches 528; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

Qy 207 GGAAGAGAAACAAAGCCATTGAGTCCCTGTCAAAGAGAGAAACCCCTG-AAAAA 265
Db 695 GGAAGAGAAACAAAGCCATTGAGTCCCTGTCAAAGAGAGAAACCCCTGAAAAA 636

Qy 266 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATTTACATCTCAATATCCAGACTG 325
Db 635 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATTTACATCTCAATATCCAGACTG 576

Qy 326 AGAGAAATGCAAGAGGGCTCAACGATTCAATGTACCTGTGAGCTTGGAGAGTAAGAAAG 385
Db 575 AGAGAAATGCAAGAGGGCTCAACGATTCAATGTACCTGTGAGCTTGGAGAGTAAGAAAG 516

Qy 386 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAAGGTCGTCACTGATA 445
Db 515 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAAGGTCGTCACTGATA 456

Qy 446 ACAAACTTATGGTTAACTTGGATAGCTGAAGAAAGAGCTCAAGATTTGGTTGATG 505
Db 455 ACAAACTTATGGTTAACTTGGATAGCTGAAGAAAGAGCTCAAGATTTGGTTGATG 396

Qy 506 TCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAAACTGAAAGAGGAGGAGCGAT 565
Db 395 TCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAAACTGAAAGAGGAGGAGCGAT 336

Qy 566 TTGGATTGTCAAGTTTCAGCTGGAACCTGGAACCAACAGAGGATACAGAGCAAGAGA 625
Db 335 TTGGATTGTCAAGTTTCAGCTGGAACCTGGAACCAACAGAGGATACAGAGCAAGAGA 276

Qy 626 GGAAGAGAGAGCGCTTTGGGATTTGGCTGATGAAAGGTTCTTGATATCTTCTGTTCTC 685
Db 626 GGAAGAGAGAGCGCTTTGGGATTTGGCTGATGAAAGGTTCTTGATATCTTCTGTTCTC 685

Db 275 GGAAGAGAGAGCGCTTTGGGATTTGGGATTTCCCTGATGAAAGTTCCTGATACTTCTGTTCTC 216
Qy 686 CAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 745
Db 215 CAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 156
Qy 746 TGCCTACGTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACCTGGG 805
Db 155 TGCCTACGTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACCTGGG 96
Qy 806 CAGCAGTTTGACTTATTTGCTTTTACGCTTTAAAGGTTTGTGTGTTT 853
Db 95 GCAGCAGTTGACTTAT-TGCTGTTTTCAGCTTTAAAGTGTGTGTTT 49

RESULT 11
US-10-006-285-233
; Sequence 233, Application US/10006285
; Publication No. US20030165854A1
; GENERAL INFORMATION:
; APPLICANT: Mary Jane Cunningham
; APPLICANT: Matthew R. Kaser
; TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS
; FILE REFERENCE: PA-0039 US
; CURRENT APPLICATION NUMBER: US/10/006,285
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 514
; SOFTWARE: PERL Program
; SEQ ID NO 233
; LENGTH: 1022
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165854A1 207354_Rn.3
; NAME/KEY: unsure
; LOCATION: 772-845
; OTHER INFORMATION: a, t, c, g, or other
US-10-006-285-233

Query Match 55.7%; Score 498.4; DB 16; Length 1022;
Best Local Similarity 87.3%; Pred. No. 3.9e-123;
Matches 580; Conservative 0; Mismatches 81; Indels 3; Gaps 3;

Qy 57 AAAGCTTTCGCAACTAAAGCAAGAATCTTGTCTGCTGCTGTTGGAGACCAAGGGAATAAA 116
Db 104 AAAGCTTTCGCAACTAAAGCAAGAATCCCTTGTCTGCGGTTTAGAGACCAAGGGAATAAA 163

Qy 117 GCAAGATCTTATCCAGACT-CCAGGCATATCTTGAAGACATGCTGAAGAGAGGACAA 175
Db 164 ACAAGATCTTATCAATAGGCTACCCAGGCATATCTTGAAGAACATGCTGAAGAGAGACAA 223

Qy 176 ATGAAGAAGATGTACTGGGAGATGAAACAGAGGAAGAAACAAAGCCATTCAGCTCC 235
Db 224 ATGAAGAAGATGTACTGGGAGATGAAACTGAGGAGAGAAACCAAGCCATATAGAACTGC 283

Qy 236 CTGTCAAAGAGGAGAAACCCCTCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGG 295
Db 284 CTGTTAAAGAGGAGAAACCTCTGAAAAAAAGTTGTTGATATGGCATCAGAAAGAGGTTG 343

Qy 296 TGAATAATACATCTGAATACCAAGACTGAGAGATGCAAGAGAGGCTGAACGATTC 355
Db 344 TAAAAATACATCTGGAATACCTCAAACTGAGAGATGCAAGAGAGGCTGAACGATTC 403

Qy 356 ATGTACCTGTGAGCTTGGAGATGAAGAAAGCTGCTCG-GGCAGCTAGGTTTGGGATTTCT 414
Db 404 ATGTGCTGTGAAGCTTGGAGATGAAGAAAGGCTGCTCGCGGAGGAGGTTTGGGATTTCT 463

Qy 415 TCAGTTCCAAACAAAGGCTGTGTCTCATCTGTATAACAAACCTATGTTTAACCTGGATAAGCTG 474
Db 464 TCAGTTCCAAACAAAGGTTTATCATCTGACACCAAGCCAAATGGTTAACTGGATAACTA 523

Qy 475 AAGAAAGAGAGCTCAAGAGATTTGGTGAATGTCTCTTCAATCTCTCCAGAAAGTCTGAAGAT 534
```


Db 421 CAAACAAAGGTCGTGTCATCTGATAACAAACCTATGGTTAACTTGGATAAG 470

Search completed: October 5, 2005, 05:55:38
Job time : 757.783 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:18:12 ; Search time 573.739 Seconds
(without alignments)
9224.147 Million cell updates/sec

Title: US-09-788-476A-1
Perfect score: 894
Sequence: 1 999agtgagtgagggttaa.....taataaaaaaaatagaaaa 894

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002as.*
7: geneseqn2002bs.*
8: geneseqn2003as.*
9: geneseqn2003bs.*
10: geneseqn2003cs.*
11: geneseqn2003ds.*
12: geneseqn2004as.*
13: geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	894	100.0	894	ABX13935	ABX13935 CDNA enco
2	894	100.0	1154	AAS29109	AAS29109 CDNA enco
3	894	100.0	1154	ABX68249	ABX68249 CDNA enco
4	894	100.0	1154	ADC25243	ADC25243 Human CDN
5	894	100.0	1520	ABN59623	ABN59623 Novel hum
6	891	99.7	1071	AAK88100	AAK88100 Human FLE
7	873	97.7	873	ABX13936	ABX13936 DNA enco
8	869.4	97.2	891	AAS62602	AAS62602 CDNA sequ
9	733.4	82.0	3147	ADP122609	ADP122609 Human liv
10	690.8	77.3	2553	ADL62461	ADL62461 Human ova
11	619	69.2	620	ACN39490	ACN39490 Tumour-as
12	595.2	66.6	2553	ADL62461	ADL62461 Human ova
13	498.4	55.7	1022	ADI22423	ADI22423 Rat liver
14	467.6	52.3	471	AAK00738	AAK00738 Human sec
15	451.6	50.5	558	ADP28791	ADP28791 Human sec
16	336.2	37.6	463	ACH28217	ACH28217 Human adu
17	334	37.4	408	AAH81555	AAH81555 Human dif
18	281.4	31.5	5469	AAK84119	AAK84119 Human imm
19	281.4	31.5	5469	AAK137635	AAK137635 Human met
20	281.4	31.5	5469	AAK06683	AAK06683 Human rep

C 21	281.4	31.5	5469	4	ABA08022	ABA08022 Human ova
C 22	281.4	31.5	5469	8	ABX60623	ABX60623 CDNA enco
C 23	281.4	31.5	5469	12	ADJ31373	ADJ31373 Human mus
C 24	281.4	31.5	9453	4	AAK84118	AAK84118 Human imm
C 25	281.4	31.5	9453	4	AAK137634	AAK137634 Human mus
C 26	281.4	31.5	9453	4	AAK06682	AAK06682 Human rep
C 27	281.4	31.5	9453	4	ABA08021	ABA08021 Human ova
C 28	281.4	31.5	9453	8	ABX60622	ABX60622 CDNA enco
C 29	281.4	31.5	9453	12	ADJ31372	ADJ31372 Human mus
C 30	258.6	28.9	266	2	AAV89379	AAV89379 EST clone
C 31	246.2	27.5	300	10	ADI22386	ADI22386 Rat liver
C 32	231.2	25.9	591	2	AAK85623	AAK85623 Novel CDN
C 33	211.8	23.7	255	5	AAH82022	AAH82022 Rat diffe
C 34	196.8	22.0	232	8	AAK49101	AAK49101 Mouse DST
C 35	164.8	18.4	399	10	ADBS0756	ADBS0756 Primary r
C 36	154.2	17.2	215	6	ABX69315	ABX69315 Novel mur
C 37	92.2	10.3	463	5	ADL41869	ADL41869 Human ova
C 38	56	6.3	208765	12	ADQ97430	ADQ97430 Mouse can
C 39	55.4	6.2	1686	2	AAQ87587	AAQ87587 DNA enco
C 40	55	6.2	399	5	AAK90682	AAK90682 DNA enco
C 41	55	6.2	399	5	AAK93406	AAK93406 DNA enco
C 42	54.4	6.1	1997	13	ADR07778	ADR07778 Full leng
C 43	54	6.0	5659	6	ABL32375	ABL32375 Human imm
C 44	54	6.0	5659	6	ABL34487	ABL34487 Human met
C 45	54	6.0	5659	7	ADS99748	ADS99748 Complem

ALIGNMENTS

RESULT 1
ABX13935
ID ABX13935 standard; cDNA; 894 BP.
XX
AC ABX13935;
XX
DT 07-MAR-2003 (first entry)
XX
DE cDNA encoding novel human protein HCC-1.
XX
KW Human; hcc-1; human hepatocellular carcinoma tissue; HCC-M cell line;
KW cancer; hepatocellular carcinoma; antisense gene therapy; gene;
KW chromosome 7q22.1; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 25..657
FT /*tag= a
FT /product= "HCC-1"
XX
XX US2002107190-A1.
XX
XX 08-AUG-2002.
XX
XX 21-FEB-2001; 2001US-00788476.
XX
XX 25-FEB-2000; 2000US-0185116P.
XX
XX (CHUN/) CHUNG C M.
XX (CHAN/) CHAN L.
XX (OUKK/) OU K.
XX (ONGS/) ONG S.
XX (SEOW/) SEOW T K.
XX (LIAN/) LIANG C R.
XX (CHOO/) CHOONG M L.
XX (TANL/) TAN L K.
XX
XX Chung CM, Chan L, Ou K, Ong S, Seow TK, Liang CR, Choong ML;
XX Tan LK;
XX
XX WPI; 2002-697878/75.
XX
XX P-PSDB; ABG72881.
XX
XX

XX New nucleic acid which is differentially expressed in human
PT hepatocellular carcinoma tissue useful for diagnosing and developing
XX PT therapy for hepatocellular carcinoma and related conditions.
PS
PS Claim 2; Fig 1; 23pp; English.
XX
XX The invention describes an isolated nucleic acid whose expression is
CC differential or preferential in human hepatocellular carcinoma tissue or
CC tissue from a related cancer relative to other tissue in the subject(s)
CC diagnosed with the condition. The nucleic acid is used to diagnose and
CC treat hepatocellular carcinoma and related cancers, or modulate one or
CC more activities in a cell e.g. by antisense gene therapy. This sequence
CC encodes the novel human protein HCC-1 identified from the HCC-W cell and
CC proposed to be involved in nucleic acid binding and transcription
CC control. The gene encoding HCC-1 is located on human chromosome 7q22.1
XX
XX SQ Sequence 894 BP; 287 A; 156 C; 233 G; 218 T; 0 U; 0 Other;

Query Match 100.0%; Score 894; DB 6; Length 894;
Best Local Similarity 100.0%; Pred. No. 2e-230;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGAGTGGAGTGAGGGGTAAACAGATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAG 60
Db 1 GGGAGTGGAGTGAGGGGTAAACAGATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAG 60

Qy 61 CTTCCGGAACCTAAAGCAAGATGCTTCTGCTCGTGGTTTGGAGACCAAGGGAATAAAGCAA 120
Db 61 CTTCCGGAACCTAAAGCAAGATGCTTCTGCTCGTGGTTTGGAGACCAAGGGAATAAAGCAA 120

Qy 121 GATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGGAGGCAATGAA 180
Db 121 GATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGGAGGCAATGAA 180

Qy 181 GAAGATGTACTGGGAGATGAAACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTC 240
Db 181 GAAGATGTACTGGGAGATGAAACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTC 240

Qy 241 AAAGAGGAAGAACCCCTGAAAAAAGTGTGATGTGGCAGAGAGAAAGAAAGTGGTGA 300
Db 241 AAAGAGGAAGAACCCCTGAAAAAAGTGTGATGTGGCAGAGAGAAAGAAAGTGGTGA 300

Qy 301 ATTACATCTGAATACACACAGACTGAGAGATGAGAGAGAGGCTGACGATTCATGTA 360
Db 301 ATTACATCTGAATACACACAGACTGAGAGATGAGAGAGAGGCTGACGATTCATGTA 360

Qy 361 CCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGATTCTTCAGTT 420
Db 361 CCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGATTCTTCAGTT 420

Qy 421 CCAACAAAGGTCTGTCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGAA 480
Db 421 CCAACAAAGGTCTGTCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGAA 480

Qy 481 AGAGCTCAAAGATTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 540
Db 481 AGAGCTCAAAGATTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 540

Qy 541 AAATCTGAAAAAGAGAGGAGCGATTGGGATTGTCAAGTTCAGCTGGAACCTGGAACC 600
Db 541 AAATCTGAAAAAGAGAGGAGCGATTGGGATTGTCAAGTTCAGCTGGAACCTGGAACC 600

Qy 601 ACAGAGGATACAGAGGCAAGAGAGGAAAAAGACAGAGCGCTTTGGGATTGCTGATGA 660
Db 601 ACAGAGGATACAGAGGCAAGAGAGGAAAAAGACAGAGCGCTTTGGGATTGCTGATGA 660

Qy 661 AAAGTTCCTGATCTTCTGTTCTCCAGTGTTCATTTCTCTCTCTCTCTCTCTCTCT 720
Db 661 AAAGTTCCTGATCTTCTGTTCTCCAGTGTTCATTTCTCTCTCTCTCTCTCTCTCT 720

Qy 721 ATATATGCTAAATGACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGCATGTA 780
Db 721 ATATATGCTAAATGACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGCATGTA 780

Db 721 ATATATGCTAAATGACAGTCATGTGCTCGCTACGTCTGCTCGCAATGAGGAGCATGTA 780

Qy 781 CCCAGGTATCATCCATGAACCTGGGCGAGAGTTTGACTTATTGCTGTTTCAGCTTTAAGG 840
Db 781 CCCAGGTATCATCCATGAACCTGGGCGAGAGTTTGACTTATTGCTGTTTCAGCTTTAAGG 840

Qy 841 TTGTTGTTTTGTTTTGTTTTGATTATGTTGCTTAAATAAAAAAATAGAAAA 894
Db 841 TTGTTGTTTTGTTTTGTTTTGATTATGTTGCTTAAATAAAAAAATAGAAAA 894

RESULT 2
AAS29109
ID AAS29109 standard; cDNA; 1154 BP.
XX
AC AAS29109;
XX
DT 21-NOV-2001 (first entry)
XX
DE cDNA encoding for human DNA-binding protein #80.
XX
KW Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; Y-box binding protein; DNA organisation;
KW gene transcription; malignant disease; autoimmune disorder;
KW rheumatic disease; genetic abnormality; infectious disease;
KW neurological disorder; gene therapy; immunomodulatory; anti-HIV;
KW anti rheumatic; anti microbial; cystostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200155162-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001305.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 29-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-023935P.
PR 13-OCT-2000; 2000US-0239357P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.

01-SEP-2000; 2000US-0249213P.
17-NOV-2000; 2000US-0249214P.
17-NOV-2000; 2000US-0249215P.
17-NOV-2000; 2000US-0249216P.
17-NOV-2000; 2000US-0249217P.
17-NOV-2000; 2000US-0249218P.
17-NOV-2000; 2000US-0249244P.
17-NOV-2000; 2000US-0249245P.
17-NOV-2000; 2000US-0249264P.
17-NOV-2000; 2000US-0249265P.
17-NOV-2000; 2000US-0249297P.
17-NOV-2000; 2000US-0249299P.
17-NOV-2000; 2000US-0249300P.
01-DEC-2000; 2000US-0250160P.
01-DEC-2000; 2000US-0250391P.
05-DEC-2000; 2000US-0251030P.
05-DEC-2000; 2000US-0251988P.
05-DEC-2000; 2000US-0256719P.
06-DEC-2000; 2000US-0251479P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
11-DEC-2000; 2000US-0254097P.
05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-465557/50.
P-PSDB; AAU18233.

Nucleic acid molecules encoding human secreted chromosomal binding proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers.
Claim 4; SEQ ID NO 90; 561pp; English.

The present invention relates to the isolation of novel DNA-binding proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding for these proteins. DNA-binding proteins such as histones, chromo (chromatin organisation modifier) domain proteins, and Y-box binding proteins may contribute to diseases resulting from aberrant DNA organisation and/or gene transcription. The sequences of the invention are useful in screening assays to identify antagonists and/or agonists that may enhance or block activities mediated by DNA-binding proteins. Blockers of DNA-binding proteins may be useful in treating disorders such as malignant diseases (e.g. cancer), autoimmune disorders (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's disease). The polynucleotide sequences of the invention may also be used in gene therapy. AAS29030-AAS29157 represent cDNA sequences encoding for novel DNA-binding proteins. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 1154 BP; 394 A; 186 C; 291 G; 281 T; 0 U; 2 Other;
Query Match 100.0%; Score 894; DB 5; Length 1154;
Best Local Similarity 100.0%; Pred. No. 2,2e-230;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 1 GGGAGTGGAGTGAGGGGTAAACAGATGGCGACCCGAGACGGTGGAGTCCTCATAGCTTAAG 60
| | | | |
Db 177 GGGAGTGGAGTGAGGGGTAAACAGATGGCGACCCGAGACGGTGGAGTCCTCATAGCTTAAG 236
| | | | |

QY 61 CTTCGCCGAACATAAGCAAGATGCTTTGCTCGTGGTTTTGGAGACCAGGGAATAAGCAA 120
| | | | |
Db 237 CTTCGCCGAACATAAGCAAGATGCTTTGCTCGTGGTTTTGGAGACCAGGGAATAAGCAA 296
| | | | |


```
XX Claim 1; SEQ ID NO 90; 225pp; English.
XX
XX The present invention relates to a new DNA-binding protein. The invention
CC is useful in treating, preventing, diagnosing and/or prognosing
CC immunodeficiencies (e.g. B cell immunodeficiencies, severe combined
CC immunodeficiencies), autoimmune disorders (rheumatoid arthritis, multiple
CC sclerosis, diabetes mellitus), allergic reactions and conditions (e.g.
CC asthma), inflammatory conditions, graft-versus-host disease, blood-
CC related disorders (thrombosis, atherosclerosis), hyperproliferative
CC disorders (e.g. cancer), renal disorders (e.g. acute glomerulonephritis),
CC cardiovascular disorders (e.g. arrhythmia), respiratory disorders
CC (Goodpasture's syndrome), neurological disorders (e.g. Alzheimer's
CC disease, Parkinson's disease), endocrine disorders (e.g. Addison's
CC disease), reproductive system disorders (e.g. endometriosis), infectious
CC diseases (e.g. viral, bacterial or fungal infections) and
CC gastrointestinal disorders (e.g. Crohn's disease). The invention is also
CC useful to stimulate neuronal growth and treat, prevent, and/or diagnose
CC neuronal damage which occurs in certain neuronal disorders or neuro-
CC degenerative conditions. The present nucleic acid sequence encodes a
CC human DNA-binding protein of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC http.seqdata.uspto.gov/sequence
XX
XX SQ Sequence 1154 BP; 394 A; 186 C; 291 G; 281 T; 0 U; 2 Other;
Query Match 100.0%; Score 894; DB 6; Length 1154;
Best Local Similarity 100.0%; Pred. No. 2.2e-230;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GGGAGTGGAGTGGAGGGTAAACAAGATGGCGACCGAGAGCGTGGAGCTCCATAAGCTAAAG 60
Db 177 GGGAGTGGAGTGGAGGGTAAACAAGATGGCGACCGAGAGCGTGGAGCTCCATAAGCTAAAG 236
Qy 61 CTTCCCGAATTAAGCAGAGATGCTTGCTCGTGGTTTGGAGACCAAGGGGAATAAAGCAA 120
Db 237 CTTCCCGAATTAAGCAGAGATGCTTGCTCGTGGTTTGGAGACCAAGGGGAATAAAGCAA 296
Qy 121 GATCTTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGGGCGCAATGAA 180
Db 297 GATCTTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGGGCGCAATGAA 356
Qy 181 GAAGATGTAATGGAGATGAACAGAGAGAAAGAAACAAAGCCCATTTAGCTCCCTGTC 240
Db 357 GAAGATGTAATGGAGATGAACAGAGAGAAAGAAACAAAGCCCATTTAGCTCCCTGTC 416
Qy 241 AAAGAGGAGAACCCCTGAAACAACTTTGATGTGGCAGCAGAGAGAAAGTGGTGA 300
Db 417 AAAGAGGAGAACCCCTGAAACAACTTTGATGTGGCAGCAGAGAGAAAGTGGTGA 476
Qy 301 ATTACATCTGAATACACAGACTGAGAGAAATGAGAGAGGGCTGAACGATTCATGTA 360
Db 477 ATTACATCTGAATACACAGACTGAGAGAAATGAGAGAGGGCTGAACGATTCATGTA 536
Qy 361 CTTGTGAGCTTGGAGAGTGAAGAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTT 420
Db 537 CTTGTGAGCTTGGAGAGTGAAGAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTT 596
Qy 421 CCAACAAAGGTCTGTCTATCTGATTAACAAACCTATGGTTAACTTGGATTAAGCTGAAGAA 480
Db 597 CCAACAAAGGTCTGTCTATCTGATTAACAAACCTATGGTTAACTTGGATTAAGCTGAAGAA 656
Qy 481 AGAGCTCAAGATTTGGTTGAATGCTCTCTCAATCTCCAGAAAGTCTGAAGATGATGAG 540
Db 657 AGAGCTCAAGATTTGGTTGAATGCTCTCTCAATCTCCAGAAAGTCTGAAGATGATGAG 716
Qy 541 AAATGAAAAAGAGAGAGAGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAACC 600
Db 717 AAATGAAAAAGAGAGAGAGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAACC 776
Qy 601 ACAGAGGATACAGAGGCAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGGCTGATGA 660
|||||
```

```
Db 777 ACAGAGATACAGAGGCAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGGCTGATGA 836
Qy 661 AAGATTCTCTGATCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
|||||
Db 837 AAGATTCTCTGATCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 896
|||||
Qy 721 ATATATGCTTAAATGACACATGTCCTACGTCCTGCTCGCAATGAGGAGCATGTA 780
|||||
Db 897 ATATATGCTTAAATGACACATGTCCTACGTCCTGCTCGCAATGAGGAGCATGTA 956
|||||
Qy 781 CCCAGGTACATCCATGAACTGGCGGACAGCTTTGACTTTGCTTTGCTTTGCTTTGCTTT 840
|||||
Db 957 CCCAGGTACATCCATGAACTGGCGGACAGCTTTGACTTTGCTTTGCTTTGCTTTGCTTT 1016
|||||
Qy 841 TTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 894
|||||
Db 1017 TTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 1070
|||||
RESULT 4
ADC25243
ID ADC25243 standard; cDNA; 1154 BP.
XX AC
XX AC ADC25243;
XX
XX 18-DEC-2003 (first entry)
XX
XX Human cDNA from extracellular matrix gene 80.
XX
XX Extracellular matrix protein; cytostatic; antibacterial; virucide;
KW neuroprotective; gynaecological; gastrointestinal-Gen; cardiant;
KW cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;
KW respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
KW nootropic; anti-allergic; cancer; bacterial infection; viral infection;
KW neural disorder; immune system disorder; blood disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW inflammatory disorder; proliferative disorder; human; gene therapy; ss;
KW gene.
XX
XX Homo sapiens.
XX
XX US2003049650-A1.
XX
XX 13-MAR-2003.
XX
XX 07-MAR-2002; 2002US-00091483.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX
XX 18-APR-2000; 2000US-0198123P.
XX
XX 19-MAY-2000; 2000US-0205151P.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX 28-JUN-2000; 2000US-0214886P.
XX
XX 30-JUN-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216647P.
XX
XX 07-JUL-2000; 2000US-0216880P.
XX
XX 11-JUL-2000; 2000US-0217487P.
XX
XX 11-JUL-2000; 2000US-0217496P.
XX
XX 14-JUL-2000; 2000US-0218290P.
XX
XX 26-JUL-2000; 2000US-0220963P.
XX
XX 26-JUL-2000; 2000US-0220964P.
XX
XX 14-AUG-2000; 2000US-0224518P.
XX
XX 14-AUG-2000; 2000US-0224519P.
XX
XX 14-AUG-2000; 2000US-0225213P.
XX
XX 14-AUG-2000; 2000US-0225214P.
XX
XX 14-AUG-2000; 2000US-0225266P.
XX
XX 14-AUG-2000; 2000US-0225267P.
XX
XX 14-AUG-2000; 2000US-0225268P.
```

PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 13-SEP-2000; 2000US-0231986P.
PR 14-SEP-2000; 2000US-0232337P.
PR 14-SEP-2000; 2000US-0232338P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.

PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764846.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM, Barash SC;

WPI; 2003-605749/57.

P-PSDB; ADC25371.

New DNA-binding proteins and gene encoding them, useful for diagnosing, treating and/or preventing e.g. neurological, inflammatory, infectious, cardiovascular, autoimmune, respiratory, neoplastic or digestive diseases.

Claim 1; SEQ ID NO 90; 226pp; English.

The invention relates to an isolated nucleic acid molecule (cDNA) encoding a human extracellular matrix protein, representing one of 161 novel genes. Also included are recombinant vectors, host cells (expressing the protein), the extracellular matrix proteins (including their fragments, epitopes and homologues), an isolated antibody that binds specifically to the protein, diagnosing a pathological condition or susceptibility to a pathological condition (comprising determining the presence or absence of a mutation in the nucleic acid and diagnosing a condition based on the presence or absence of the mutation), diagnosing a pathological condition or susceptibility to a pathological condition (comprising determining the presence or amount of expression of the protein in a biological sample and diagnosing a condition based on the presence or amount of expression of the protein), preventing, treating or ameliorating a medical condition by administering the nucleic acid or protein to a mammalian subject, identifying a binding partner to the protein, the gene corresponding to the cDNA sequence, and identifying an activity in a biological assay (comprising expressing the nucleic acid in a cell, isolating the supernatant, detecting an activity in a biological assay and identifying the protein in the supernatant having the

CC activity). The nucleic acids and proteins display the following
CC activities Cytostatic, antibacterial, Virucide, Neuroprotective,

Query Match 100.0%; Score 894; DB 10; Length 1154;
Best Local Similarity 100.0%; Pred. No. 2.2e-230;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGAGTGCAGTGGGGTAAACAAGATGCGGACCGAGAGCGTGGAGCTCCATAAGCTAAAG 60
Db 177 GGGAGTGGAGTGAGGGGTAAACAAGATGCGGACCGAGAGCGTGGAGCTCCATAAGCTAAAG 236
QY 61 CTTGCCGAACATAAGCAAGATGCTTCTGCTCGTGGTTTGGAGACCAAGGGAATAAAGCAA 120
Db 237 CTTGCCGAACATAAGCAAGATGCTTCTGCTCGTGGTTTGGAGACCAAGGGAATAAAGCAA 296
QY 121 GATCTTATCCACAGACTCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAAAATGAA 180
Db 297 GATCTTATCCACAGACTCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAAAATGAA 356
QY 181 GAAGATGCTACTGGGAGTGAACAGAGCAAGAGAAAGAAAGCCATTTGAGCTCCCTGTC 240
Db 357 GAAGATGCTACTGGGAGTGAACAGAGCAAGAGAAAGAAAGCCATTTGAGCTCCCTGTC 416
QY 241 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTGA 300
Db 417 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTGA 476
QY 301 ATTACATCTGAAATACCAAGACTGAGAGAAATGAGAAAGGGCTGGAACGATTCAATGTA 360
Db 477 ATTACATCTGAAATACCAAGACTGAGAGAAATGAGAAAGGGCTGGAACGATTCAATGTA 536
QY 361 CTTGAGCTTGGAGAGTGAAGAGCTGCTCGGAGCTAGCTTTGGGATTTCTTCAGTT 420
Db 537 CTTGAGCTTGGAGAGTGAAGAGCTGCTCGGAGCTAGCTTTGGGATTTCTTCAGTT 596
QY 421 CCAACAAAGTCTGTCTATCTGATAACAACTATGTTAACTTGGATTAAGCTGAAGGAA 480
Db 597 CCAACAAAGTCTGTCTATCTGATAACAACTATGTTAACTTGGATTAAGCTGAAGGAA 656
QY 481 AGAGCTCAAGATTTGGTTGAATGCTCTCTCAATCTCAGAAAGTCTGAAGATGATGAG 540
Db 657 AGAGCTCAAGATTTGGTTGAATGCTCTCTCAATCTCAGAAAGTCTGAAGATGATGAG 716
QY 541 AAATCTGAAAAGAGAGGAGCGATTTGGGATTTGTCAAGTTCAGCTGGAACCTGGAAC 600
Db 717 AAATCTGAAAAGAGAGGAGCGATTTGGGATTTGTCAAGTTCAGCTGGAACCTGGAAC 776
QY 601 ACAGAGGATACAGAGGCAAGAGGAAAGAGAGAGCGCTTTGGGATTTGCTGATGA 660
Db 777 ACAGAGGATACAGAGGCAAGAGGAAAGAGAGAGCGCTTTGGGATTTGCTGATGA 836
QY 661 AAAGTTCCTGATCTTTCTGTTCTCCAGTGTTCCTTCTCTCTCTCTCTCTCTCTCTCT 720
Db 837 AAAGTTCCTGATCTTTCTGTTCTCCAGTGTTCCTTCTCTCTCTCTCTCTCTCTCTCT 896
QY 721 ATATATGCTAAATGACAGTCAATGCTGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 897 ATATATGCTAAATGACAGTCAATGCTGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCT 956
QY 781 CCCAGGTACATCCATGAACTGCGGAGCAGCTTTGACTTATGCTGTTTACGCTTTAAGG 840
Db 957 CCCAGGTACATCCATGAACTGCGGAGCAGCTTTGACTTATGCTGTTTACGCTTTAAGG 1016
QY 841 TTGTTGCTTTTGTGTTTGTATGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 894
Db 1017 TTGTTGCTTTTGTGTTTGTATGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1070

RESULT 5
ABN59623
ID ABN59623 standard; cDNA; 1520 BP.
XX
AC ABN59623;

XX 28-JUN-2002 (first entry)
XX Novel human coding sequence SEQ ID NO: 34.
XX Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
XX antifertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
XX neuroprotective; antiparkinsonian; protein therapy; EST;
XX expressed sequence tag; gene; ss.
OS Homo sapiens.
XX WO200222660-A2.
XX 21-MAR-2002.
XX 10-SEP-2001; 2001WO-US026015.
XX 11-SEP-2000; 2000US-00659671.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX P-PSDB; ABB97210.
XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis.
XX Claim 1; SEQ ID NO 34; 509pp; English.
XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibit e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
XX and cancer, to screen for drugs, to treat inflammatory conditions e.g.
XX rheumatoid arthritis, and to treat nervous system disorders e.g.
XX parkinson's disease. The present sequence is a coding sequence of the
XX invention
XX
SQ Sequence 1520 BP; 442 A; 310 C; 423 G; 345 T; 0 U; 0 Other;
Query Match 100.0%; Score 894; DB 6; Length 1520;
Best Local Similarity 100.0%; Pred. No. 2.5e-230;
Matches 894; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGAGTGCAGTGGGGTAAACAAGATGCGGACCGAGAGCGTGGAGCTCCATAAGCTAAAG 60
Db 612 GGGAGTGCAGTGGGGTAAACAAGATGCGGACCGAGAGCGTGGAGCTCCATAAGCTAAAG 671
QY 61 CTTGCCGAACATAAGCAAGATGCTTCTGCTCGTGGTTTGGAGACCAAGGGAATAAAGCAA 120
Db 672 CTTGCCGAACATAAGCAAGATGCTTCTGCTCGTGGTTTGGAGACCAAGGGAATAAAGCAA 731
QY 121 GATCTTATCCACAGACTCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAAAATGAA 180
Db 732 GATCTTATCCACAGACTCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAAAATGAA 791
QY 181 GAAGATGCTACTGGGAGATGAAACAGAGCAAGAGAAAGAAAGCCATTTGAGCTCCCTGTC 240
Db 792 GAAGATGCTACTGGGAGATGAAACAGAGCAAGAGAAAGAAAGCCATTTGAGCTCCCTGTC 851
QY 241 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTGA 300
Db 852 AAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTGA 911
QY 301 ATTACATCTGAAATACCAAGACTGAGAGAAATGAGAAAGGGCTGGAACGATTCAATGTA 360

Qy	484	GCTCAAAGATTGGTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAA	543
Db	643	GCTCAAAGATTGGTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAA	702
Qy	544	CTGAAAAAGAGAGGAGGACGATTGGGATGCTCAACAAGTTCCAGCTGGAACTTGGAAACCACA	603
Db	703	CTGAAAAAGAGAGGAGGACGATTTGGGATGCTCAACAAGTTCCAGCTGGAACTGGAACCA	762
Qy	604	GAGGATACAGAGGCCAAAGAGAGAAAGAGCAGAGCGCTTTGGGATGCTCGATGAAAA	663
Db	763	GAGGATACAGAGGCCAAAGAGAGAAAGAGCAGAGCGCTTTGGGATGCTCGATGAAAA	822
Qy	664	GTTCTCGATACTTTCTGTTTCTCCAGTGTTTTCACATTTCTCTCCTTCTTCTGGTCAATA	723
Db	823	GTTCTCGATACTTTCTGTTTCTCCAGTGTTTTCACATTTCTCTCCTTCTTCTGGTCAATA	882
Qy	724	TATSCCTAAATGCACAGTCATGTCGCTACGTCGCTCGCAATGAGGGAGCATGTACCC	783
Db	883	TATGCTTAAATGCAACAGTCATGTCGCTACGTCGCTCGCAATGAGGGAGCATGTACCC	942
Qy	784	CAGGTACATCCATCAAATCGCGGCACAGTTTGACTTATTTGCTGTTTTCAGCTTTAAGGTTG	843
Db	943	CAGGTACATCCATCAAATCGCGGCACAGTTTGACTTATTTGCTGTTTTCAGCTTTAAGGTTG	1002
Qy	844	TTTGTTTTTTGTTTTGATTAAGTTGCTGTGTTAATAAAAAAATAGAAAA	894
Db	1003	TTTGTTTTTTGTTTTGATTAAGTTGCTGTGTTAATAAAAAAATAGAAAA	1053

PT	therapy for hepatocellular carcinoma and related conditions.
XX	Claim 2; Fig 3; 23pp; English.
XX	The invention describes an isolated nucleic acid whose expression is
CC	differential or preferential in human hepatocellular carcinoma tissue or
CC	tissue from a related cancer relative to other tissue in the subject(s)
CC	diagnosed with the condition. The nucleic acid is used to diagnose and
CC	treat hepatocellular carcinoma and related cancers, or modulate one or
CC	more activities in a cell e.g. by antisense gene therapy. This sequence
CC	represents a PCR extended form of the cDNA encoding the novel human
CC	protein HCC-1 identified from the HCC-M cell line, used in the creation
CC	of a HCC-1 expression vector
XX	
SQ	Sequence 873 BP; 272 A; 156 C; 228 G; 217 T; 0 U; 0 Other;
Query Match 97.7%; Score 873; DB 6; Length 873;	
Best Local Similarity 100.0%; Pred. No. 9e-225;	
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	6 TGGAGTGGGGGTAACAAGATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 65
Db	1 TGGAGTGGGGGTAACAAGATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 60
Qy	66 CGAACTAAACCAAGATGCTTCTGCTGTGTTGGAGACCAAGGGATAAAGCAAGTCT 125
Db	61 CGAACTAAACCAAGATGCTTCTGCTGTGTTGGAGACCAAGGGATAAAGCAAGTCT 120
Qy	126 TATCCACAGACTCCAGGCATATCTGAAGAAACATGCTGAAGAGGAGGCAANTGAAGA 185
Db	121 TATCCACAGACTCCAGGCATATCTGAAGAAACATGCTGAAGAGGAGGCAANTGAAGA 180
Qy	186 TGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTGAGCTCCCTCTCAAAGA 245
Db	181 TGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTGAGCTCCCTCTCAAAGA 240
Qy	246 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAATTAC 305
Db	241 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGTGAAATTAC 300
Qy	306 ATCTGAAATACACAGACTCAGAGATGCAAGAGAGGGCTGAACGATTCATGTACCTGT 365
Db	301 ATCTGAAATACACAGACTCAGAGATGCAAGAGAGGGCTGAACGATTCATGTACCTGT 360
Qy	366 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTCCAAC 425
Db	361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTCCAAC 420
Qy	426 AAAAGTCTGTCTATCTGATAAACAACTATGGTTAACTTGGATTAAGCTGAAGAAAGAGC 485
Db	421 AAAAGTCTGTCTATCTGATAAACAACTATGGTTAACTTGGATTAAGCTGAAGAAAGAGC 480
Qy	486 TCAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATCAGAAACT 545
Db	481 TCAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATCAGAAACT 540
Qy	546 GAAAAAGAGAAAGAGCGATTGGGATTTGTCACAAAGTTTCAGCTGGAACTGGAAACACAGA 605
Db	541 GAAAAAGAGAAAGAGCGATTGGGATTTGTCACAAAGTTTCAGCTGGAACTGGAAACACAGA 600
Qy	606 GGATACAGAGCCAAAGAGAGGAAAAAGACAGAGCGCTTTGGGATTTGCCTGTATGAAAAGT 665
Db	601 GGATACAGAGCCAAAGAGAGGAAAAAGACAGAGCGCTTTGGGATTTGCCTGTATGAAAAGT 660
Qy	666 TCCTGATACCTTCTGTTCTCCAGTGTTTCCATTTCTCTCTTCTTCTTGGTTCACATATA 725
Db	661 TCCTGATACCTTCTGTTCTCCAGTGTTTCCATTTCTCTCTTCTTCTTGGTTCACATATA 720
Qy	726 TGCTTAATGCACAGTCATGCTAGCTCCTCGCCTCGCAATCAGGAGCATGTACCCCA 785
Db	721 TGCTTAATGCACAGTCATGCTAGCTCCTCGCCTCGCAATCAGGAGCATGTACCCCA 780
Qy	786 GGTACATCCATGAATCGCGCAGCAGTTTGACTTATTTGCTGTTTTCAGCTTTAAGTTGTT 845

Db	781	GGTACATCATGAACCTGGCGACGAGTTTGACCTATTGCTGTTTTCAGCTTTAAGGTTGTT	840
Qy	846	GTGTTTTGTTTGTGATTAATGTTGCTTGTAAAT	878
Db	841	GTGTTTTGTTTGTGATTAATGTTGCTTGTAAAT	873
RESULT 8			
AA	AAS62602/c		
ID	AAS62602 standard; cDNA; 891 BP.		
XX	AAS62602;		
XX	14-FEB-2002 (first entry)		
XX	cDNA sequence #389 encoding novel human secreted protein.		
XX	Human secreted protein; hyperproliferative disorder; autoimmune disorder;		
KW	immune deficiency disorder; blood disorder; inflammatory disorder;		
KW	infectious disorder; gene therapy; antimicrobial; hepatotropic;		
KW	immunosuppressive; antirheumatic; ss.		
XX	Homo sapiens.		
OS	WO200177291-A2.		
XX	18-OCT-2001.		
XX	29-MAR-2001; 2001WO-US010485.		
XX	06-APR-2000; 2000US-0195604P.		
XX	(GEMY) GENETICS INST INC.		
XX	Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;		
PI	Gulukota K, Graham JR;		
XX	WPI; 2002-010900/01.		
DR	New polynucleotides encoding secreted proteins useful for treating e.g.		
PT	asthma, HIV and Crohn's disease.		
XX	Claim 1; Page 280; 391pp; English.		
XX	The present invention relates to the isolation of novel cDNA sequences		
CC	which encode human secreted proteins. The cDNA sequences have been		
CC	derived from a variety of human tissues. The invention also provides a		
CC	method for producing proteins from these polynucleotide sequences. The		
CC	proteins are useful for identifying compounds that modulate their		
CC	activity and production, and the cell is also useful for identifying		
CC	compounds that modulate expression of the polynucleotide sequences		
CC	encoding the secreted proteins. The sequences of the invention are useful		
CC	for treating diseases such as hyperproliferative disorders (e.g. cancer),		
CC	immune deficiency disorders (e.g. severe combined immunodeficiency		
CC	(SCID), autoimmune disorders (e.g. multiple sclerosis), blood disorders		
CC	(e.g. thrombocytopaenia), inflammatory disorders (e.g. arthritis) and		
CC	infectious disorders (e.g. hepatitis). The polynucleotide sequences of		
CC	the invention are also useful in gene therapy. AAS62214-AAS62838		
CC	represent the cDNA sequences of the invention that encode for novel human		
CC	secreted proteins		
XX	Sequence 891 BP; 221 A; 224 C; 157 G; 289 T; 0 U; 0 Other;		
SQ	Query Match 97.2%; Score 869.4; DB 6; Length 891;		
	Best Local Similarity 99.8%; Pred. No. 8.5e-224;		
	Matches 881; Conservative 0; Mismatches 1; Indels 1; Gaps 1;		
Qy	13	AGGGTTAAAGATGGCGACGACGCGTCCATAGCTTCAATAGCTTGGCGAACTA	72
Db	886	AAGGGTAAAGATGGCGACGACGCGTCCATAGCTTCAATAGCTTGGCGAACTA	827
Qy	73	AAGCAAGAATGTCTTGCTCGTGGTTTGAGACCAAGGGAATAAGCAAGATCTTATCCAC	132

Db	826	AAGCAAGAATGTCTTGCTCGTGGTTTGAGACCAAGGGAATAAGCAAGATCTTATCCAC	767
Qy	133	AGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAATAAGCAAGATGTACTG	132
Db	766	AGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAATAAGCAAGATGTACTG	707
Qy	193	GGAGATGAAACAGAGGAAGAAGAAACAAAGCCATTTGAGCTCCCTGTCAAAGAGGAAGAA	252
Db	706	GGAGATGAAACAGAGGAAGAAGAAACAAAGCCATTTGAGCTCCCTGTCAAAGAGGAAGAA	647
Qy	253	CCCCCTGAAAAAAGCTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTGAATAATTACATCTGAA	312
Db	646	CCCCCTGAAAAAAGCTGTTGATGTGGCAGCAGAGAAAGAAAGTGGTGAATAATTACATCTGAA	587
Qy	313	ATACCACAGACTGAGAGAAATGAGAGAGGCTGAAACGATTCATCTACTCTGTGAGCTTG	372
Db	586	ATACCACAGACTGAGAGAAATGAGAGAGGCTGAAACGATTCATCTACTCTGTGAGCTTG	527
Qy	373	GAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTCACAAAGAGGT	432
Db	526	GAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTCACAAAGAGGT	467
Qy	433	CTGTCACTGTATAACAAACCTATGTTAACTTTGGATAAGCTGAAGGAAAGAGCTCAAAGA	492
Db	466	CTGTCACTGTATAACAAACCTATGTTAACTTTGGATAAGCTGAAGGAAAGAGCTCAAAGA	407
Qy	493	TTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAGAG	552
Db	406	TTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAGAG	347
Qy	553	AGGAAGGAGCGATTTGGGATTTGTCAACAGTTTCAGCTGGAAAC-TGGAAACCAAGAGGATAC	611
Db	346	AGGAAGGAGCGATTTGGGATTTGTCAACAGTTTCAGCTGGAACTTGGAACTCCAGAGGATAC	287
Qy	612	AGAGGCAAAAGAGAGAAAGAGAGAGCGCTTTGGGATTTGGCTGTGATGAAAGTTCTCTGA	671
Db	286	AGAGGCAAAAGAGAGAGAAAGAGAGAGCGCTTTGGGATTTGGCTGTGATGAAAGTTCTCTGA	227
Qy	672	TACTTCTCTGTTCCAGTGTGTTTCCAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTA	731
Db	226	TACTTCTCTGTTCCAGTGTGTTTCCAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTA	167
Qy	732	AATGCACAGTCATGTCCTACGTCCTCGCAATGAGGAGCAGTGTACCCAGGTACA	791
Db	166	AATGCACAGTCATGTCCTACGTCCTCGCAATGAGGAGCAGTGTACCCAGGTACA	107
Qy	792	TCCATCAACTGCGGCGAGCAGTTTGACTTATTTGCTGTTTTCAGCTTTAAGGTTTGTGTTT	851
Db	106	TCCATCAACTGCGGCGAGCAGTTTGACTTATTTGCTGTTTTCAGCTTTAAGGTTTGTGTTT	47
Qy	852	TTGTTTTTCAATATGTTGCTTGTGTTTAAAAAATAAGAAAA	894
Db	46	TTGTTTTTCAATATGTTGCTTGTGTTTAAAAAATAAGAAAA	4
RESULT 9			
AD	ADI22609		
ID	ADI22609 standard; DNA; 3147 BP.		
XX	ADI22609;		
XX	22-APR-2004 (first entry)		
XX	Human liver differentially expressed cDNA seq id 419.		
XX	hepatotropic; antiinflammatory; virucide; cytostatic; gene therapy;		
KW	differentially gene expression; liver; toxin; liver disorder;		
KW	biliary cirrhosis; X-linked adrenoleukodystrophy; Zellweger syndrome;		
KW	hepatorenal syndrome; hepatitis; hepatocarcinoma; metabolic response;		
KW	toxicological response; ss; EST; expressed sequence tag; human.		
XX	Homo sapiens.		
OS			

XX US0003165854-A1.
XX 04-SEP-2003.
XX 05-DEC-2001; 2001US-00006285.
XX 05-DEC-2000; 2000US-0251986P.
XX (CUNN/) CUNNINGHAM M J.
XX (KASE/) KASER M R.
XX Cunningham MJ, Kaser MR;
XX WPI; 2003-863697/80.
XX
XX New combination comprising a number of cDNAs that are differentially
XX expressed in a liver treated with a toxin, useful for diagnosing, staging
XX or treating liver disorders (e.g. biliary cirrhosis, hepatitis or
XX hepatocarcinoma).
XX
XX Claim 1; SEQ ID NO 419; 28pp; English.
XX
XX The invention describes a combination comprising a number of cDNAs that
XX are differentially expressed in a liver treated with a toxin and are
XX selected from any of the 514 cDNAs listed in the specification, or their
XX complements. The combination is useful in diagnosing, staging or treating
XX a liver disorder (e.g. biliary cirrhosis, X-linked adrenoleukodystrophy,
XX Zellweger syndrome, hepatorenal syndrome, hepatitis or hepatocarcinoma),
XX in monitoring diagnostic and therapeutic applications, in detecting
XX metabolic and toxicological responses, and in elucidating drug mechanism
XX of action. This sequence represents a cDNA differentially expressed in
XX liver tissues in response to treatment with a toxin.
XX
XX SQ Sequence 3147 BP; 809 A; 757 C; 756 G; 825 T; 0 U; 0 Other;

Query Match 82.0%; Score 733.4; DB 10; Length 3147;
Best Local Similarity 99.9%; Pred. No. 6.7e-187;
Matches 734; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 57 AAGCTTGCCAACTAAGCAAGATGCTTGTCTGCTGTTGGAGCAAGGGAATAA 116
DB 97 AAGCTTGCCAACTAAGCAAGATGCTTGTCTGCTGTTGGAGCAAGGGAATAA 156
QY 117 GCAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAA 176
DB 157 GCAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAA 216
QY 177 TGAAGAGATGTACTGGAGATGAAACAGAGGAAGAGAAACAAAGCCCATTTAGCTCCC 236
DB 217 TGAAGAGATGTACTGGAGATGAAACAGAGGAAGAGAAACAAAGCCCATTTAGCTCCC 276
QY 237 TGTCAAGAGAGAACCCCTGAAACAACTGTTGATGTGCGCAGCAGAGAGAAAGTGGT 296
DB 277 TGTCAAGAGAGAACCCCTGAAACAACTGTTGATGTGCGCAGCAGAGAGAAAGTGGT 336
QY 297 GAAATTAATCATCTGAAATACCAACAGACTGAGAGATGCAAGAGAGGCTGAACGATTCAA 356
DB 337 GAAATTAATCATCTGAAATACCAACAGACTGAGAGATGCAAGAGAGGCTGAACGATTCAA 396
QY 357 TGTACCTGTAGCTTGGAGATGAAAGAGCTGCTGGCAGCTAGTTGGATTTCTTC 416
DB 397 TGTACCTGTAGCTTGGAGATGAAAGAGCTGCTGGCAGCTAGTTGGATTTCTTC 456
QY 417 AGTTCCAAACAAAGTCTGTATCTGATGATAACAAACCTATGTTAACTTGGTAAGCTCAA 476
DB 457 AGTTCCAAACAAAGTCTGTATCTGATGATAACAAACCTATGTTAACTTGGTAAGCTCAA 516
QY 477 GGAAGAGCTCAAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGA 536
DB 517 GGAAGAGCTCAAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGA 576
QY 537 TGAGAACTGAAAAAGAGAGGAGCGATTGGGATTGTCAAGTTTCAGCTGGAAGTGG 596

Db 577 TGAGAACTGAAAAAGAGAGGAGCGATTGGGATTGTCAAGTTCAAGTGAAGTGG 636
QY 597 AACCACAGAGATACAGAGGCAAAAGAGAGAGAGAGAGCGCTTTGGGATTGCTG 656
Db 637 AACCACAGAGATACAGAGGCAAAAGAGAGAGAGAGAGCGCTTTGGGATTGCTG 696
QY 657 ATGAAAAGTTCTGATATCTTCTGTTCTCCAGTCTTTTCCATTTCTCTCTCTTCTG 716
Db 697 ATGAAAAGTTCTGATATCTTCTGTTCTCCAGTCTTTTCCATTTCTCTCTCTTCTG 756
QY 717 TCACATATATGCTTAAATGCACAGTCAATGTGCTTACCTGCTCGCAATGAGGAGCA 776
Db 757 TCACATATATGCTTAAATGCACAGTCAATGTGCTTACCTGCTCGCAATGAGGAGCA 816
QY 777 TGTACCCCAAGGTACA 791
Db 817 TGTACCCCAAGTTC 831

RESULT 10
ADL62461
ID ADL62461 standard; DNA; 2553 BP.
XX AC ADL62461;
XX AC ADL62461;
DT 20-MAY-2004 (first entry)
XX Human ovarian cancer DNA marker #20673.
XX Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX Homo sapiens.
OS WO200170979-A2.
FN 27-SEP-2001.
PD 21-MAR-2001; 2001WO-US009126.
PF 21-MAR-2000; 2000US-0191031P.
PR 25-MAY-2000; 2000US-0207124P.
PR 15-JUN-2000; 2000US-0211940P.
PR 07-JUL-2000; 2000US-0216820P.
PR 25-JUL-2000; 2000US-0220661P.
PR 21-DEC-2000; 2000US-0257672P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Lee J, Lillie J;
XX WPI; 2001-611502/70.
XX
XX Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
XX cancer cells as compared to their normal non-cancerous ovarian cells are
XX used to characterize stage, grade, histological type of ovarian cancer.
XX
XX Disclosure; SEQ ID NO 20673; 106pp; English.
XX
XX The invention relates to nucleic acid markers which are overexpressed in
XX ovarian cancer cells as compared to their expression in normal (i.e. non-
XX cancerous) ovarian cells. The invention also relates to polypeptides
XX encoded by the markers, antibodies that selectively bind to the
XX polypeptides, a method of inhibiting ovarian cancer in a patient at risk
XX of developing ovarian cancer involving inhibiting expression of a gene
XX corresponding to a marker of the invention and a method of treating a
XX patient afflicted with ovarian cancer comprising providing to cells of
XX the patient an antisense oligonucleotide complementary to a marker of the
XX invention. The markers are useful for assessing if a patient is afflicted
XX with ovarian cancer, which involves comparing the level of expression of
XX a marker in a patient sample and a normal level of expression of the
XX marker in a control non-ovarian cancer sample. A difference between the
XX expression levels indicates ovarian cancer. The level of expression of a

Db 335 TTGGGATGTCACAGTTTCAGCTGGAACTGGAACCCAGAGGATACAGAGGCAAGAGA 276
QY 626 GGAAGAGAGAGAGCGCTTTGGGATTCCTGATGAAAGTTCCTGATCTTCTGTTCTC 685
Db 275 GGAAGAGAGAGAGCGCTTTGGGATTCCTGATGAAAGTTCCTGATCTTCTGTTCTC 216
QY 686 CAGTGTTTCCATTTCT 745
Db 215 CAGTGTTTCCATTTCT 156
QY 746 TGCTACGTCCTGCTCGCAATCAGGAGCATGTACCCAGGTACATCCATCAACTGCGG 805
Db 155 TGCTACGTCCTGCTCGCAATCAGGAGCATGTACCCAGGTACATCCATCAACTGCGG 96
QY 806 CAGCAGTTTGACTTATGCTGTTTTCAGCTTTAAGGTTGTTGTTT 853
Db 95 GCAGCAGTTTGACTTA-TGCTGTTTTCAGCTTTAAGGTTGTTGTTT 49

RESULT 13

AD122423
ID AD122423 standard; DNA; 1022 BP.

XX AC AD122423;

XX DT 22-APR-2004 (first entry)

XX Rat liver differentially expressed cDNA seq id 233.

XX hepatotropic; antiinflammatory; virucide; cytostatic; gene therapy;
XX differentially gene expression; liver; toxin; liver disorder;
XX biliary cirrhosis; x-linked adrenoleukodystrophy; Zellweger syndrome;
XX hepatorenal syndrome; hepatitis; hepatocarcinoma; metabolic response;
XX toxicological response; ss; EST; expressed sequence tag; rat.

XX OS Rattus sp.

XX US2003165854-A1.

XX 04-SEP-2003.

XX 05-DEC-2001; 2001US-00006285.

XX 05-DEC-2000; 2000US-0251986P.

XX (CUNN/) CUNNINGHAM M J.

XX (KASE/) KASER M R.

XX Cunnigham MJ, Kaser MR;

XX WPI; 2003-863697/80.

XX New combination comprising a number of cDNAs that are differentially
XX expressed in a liver treated with a toxin, useful for diagnosing, staging
XX or treating liver disorders (e.g. biliary cirrhosis, hepatitis or
XX hepatocarcinoma).

XX Claim 1; SEQ ID NO 233; 28pp; English.

XX The invention describes a combination comprising a number of cDNAs that
XX are differentially expressed in a liver treated with a toxin and are
XX selected from any of the 514 cDNAs listed in the specification, or their
XX complements. The combination is useful in diagnosing, staging or treating
XX a liver disorder (e.g. biliary cirrhosis, x-linked adrenoleukodystrophy,
XX Zellweger syndrome, hepatorenal syndrome, hepatitis or hepatocarcinoma),
XX in monitoring diagnostic and therapeutic applications, in detecting
XX metabolic and toxicological responses, and in elucidating drug mechanism
XX of action. This sequence represents a cDNA differentially expressed in
XX liver tissues in response to treatment with a toxin.

XX Sequence 1022 BP; 309 A; 160 C; 229 G; 250 T; 0 U; 74 Other;

Query Match 55.7%; Score 498.4; DB 10; Length 1022;
Best Local Similarity 87.3%; Pred. No. 9.2e-124;
Matches 580; Conservative 0; Mismatches 81; Indels 3; Gaps 3;
QY 57 AAAGCTTGCAGCACTAAAGCAAGAAATGTTCTGCTCGTGGTGGAGACCAAGGGAATAAA 116
Db 104 AAAGCTTGCAGCACTAAAGCAAGAAATGTTCTGCTCGGTTTAGAGACCAAGGGAATAAA 163
QY 117 GCAGAGTCTTATCCACAGACT-CCAGGCATATCTTGAAGAACATGCTGAGAGGAGGCA 175
Db 164 ACAAGATCTTATCAATAGGCTACCGGCATATCTTGAAGAACATGCTGAGAGGAGGCA 223
QY 176 ATGAAGAGAGTCTACTTGGGAGATGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 235
Db 224 ATGAAGAGAGTCTACTTGGGAGATGAAACATGAGAGAGAGAGAGAGAGAGAGAGAG 283
QY 236 CTGTCAAAG 295
Db 284 CTGTCAAAG 343
QY 296 TGAAAGATCTGAAATACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 355
Db 344 TAAAGATCTGAAATACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
QY 356 ATGTACCTGTGAGCTTGGAGAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 414
Db 404 ATGTGCTGTGAGCTTGGAGAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 463
QY 415 TCAGTTCCAAAG 474
Db 464 TCAGTTCCAAAG 523
QY 475 AGGAG 534
Db 524 AAGGAG 583
QY 535 GATGAG 594
Db 584 GATGAG 643
QY 595 GGAACACAG 654
Db 644 GGAACACAG 703
QY 655 TGATGAG 713
Db 704 TAATGAG 763
QY 714 TGGT 717
Db 764 GGT 767

RESULT 14

AAC00738
ID AAC00738 standard; cDNA; 471 BP.

XX AC AAC00738;

XX DT 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 736.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-00200610.

XX PR 26-FEB-1999; 99US-0122487P.
XX PA (GEST) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-500381/45.
XX DR P-PSDB; AAG00732.
XX PR New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX PS Claim 1; SEQ ID NO 736; 71pp + Sequence Listing; English.
XX CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors
XX SQ Sequence 471 BP; 165 A; 81 C; 130 G; 94 T; 0 U; 1 Other;

Query Match 52.3%; Score 467.6; DB 3; Length 471;
Best Local Similarity 99.8%; Pred. No. 1.3e-115;
Matches 467; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 AGTGGAGTGGGGTAAACAGATGGGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTT 63
DB 4 AGTGGAGTGGGGTAAACAGATGGGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTT 63

QY 64 GCCGAATTAAGCAAGATGCTTCTCGTGGTGGGACCAAGCAAGCAAGAT 123
DB 64 GCCGAATTAAGCAAGATGCTTCTCGTGGTGGGACCAAGCAAGCAAGAT 123

QY 124 CTTATCCACAGACTCCAGGCAATCTTGAAGCAATCTGAAGAGGAGGCAAAATGAAGAA 183
DB 124 CTTATCCACAGACTCCAGGCAATCTTGAAGCAATCTGAAGAGGAGGCAAAATGAAGAA 183

QY 184 GATGTACTGGAGATGAACACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTCAAA 243
DB 184 GATGTACTGGAGATGAACACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTCAAA 243

QY 244 GAGGAAGACCCCTGAAAACCTTTGATGTGCAGCAGAGAAAGTGGTGAATTT 303
DB 244 GAGGAAGACCCCTGAAAACCTTTGATGTGCAGCAGAGAAAGTGGTGAATTT 303

QY 304 ACATCTGAATACACAGACTGAGAGATGTCAGAGAGGGCTGAAACGATTCAATGTACCT 363
DB 304 ACATCTGAATACACAGACTGAGAGATGTCAGAGAGGGCTGAAACGATTCAATGTACCT 363

QY 364 GTGAGCTTGGAGAGTAAGAAGCTGCTCGGACAGCTAGGTTTGGGATTTCTTCAGTTCCA 423
DB 364 GTGAGCTTGGAGAGTAAGAAGCTGCTCGGACAGCTAGGTTTGGGATTTCTTCAGTTCCA 423

QY 424 ACAAAAGGTCTGTCATCTGTATAACAAACCTATGTTTAACTTTGGATAAG 471
DB 424 ACAAAAGGTCTGTCATCTGTATAACAAACCTATGTTTAACTTTGGATAAG 471

RESULT 15
ADP28791
ID ADP28791 standard; DNA; 558 BP.

XX AC ADP28791;
XX DT 12-AUG-2004 (first entry)
XX DE Human secreted protein encoding sequence SEQ ID #789.
XX KW Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;
XX KW cancer; inflammatory; immune; ds; human secreted protein.
XX OS Homo sapiens.
XX PN WO2004035732-A2.
XX PD 29-APR-2004.
XX PF 28-AUG-2003; 2003WO-US026780.
XX PR 29-AUG-2002; 2002US-0406576P.
XX PR 29-AUG-2002; 2002US-0406579P.
XX PR 29-AUG-2002; 2002US-0406585P.
XX PR 29-AUG-2002; 2002US-0406588P.
XX PR 29-AUG-2002; 2002US-0406608P.
XX PR 29-AUG-2002; 2002US-0406611P.
XX PR 29-AUG-2002; 2002US-0406612P.
XX PR 29-AUG-2002; 2002US-0406616P.
XX PR 29-AUG-2002; 2002US-0406640P.
XX PR 29-AUG-2002; 2002US-0406642P.
XX PR 29-AUG-2002; 2002US-0406646P.
XX PR 29-AUG-2002; 2002US-0406653P.
XX PR 29-AUG-2002; 2002US-0406655P.
XX PR 29-AUG-2002; 2002US-0406666P.
XX PR 17-SEP-2002; 2002US-0410946P.
XX PR 17-SEP-2002; 2002US-0410947P.
XX PR 17-SEP-2002; 2002US-0410948P.
XX PR 17-SEP-2002; 2002US-0410949P.
XX PR 17-SEP-2002; 2002US-0410953P.
XX PR 17-SEP-2002; 2002US-0410957P.
XX PR 17-SEP-2002; 2002US-0410958P.
XX PR 17-SEP-2002; 2002US-0410959P.
XX PR 17-SEP-2002; 2002US-0410960P.
XX PR 17-SEP-2002; 2002US-0410961P.
XX PR 17-SEP-2002; 2002US-0410962P.
XX PR 17-SEP-2002; 2002US-0411019P.
XX PR 17-SEP-2002; 2002US-0411022P.
XX PR 17-SEP-2002; 2002US-0411023P.
XX PR 17-SEP-2002; 2002US-0411024P.
XX PR 17-SEP-2002; 2002US-0411032P.
XX PR 17-SEP-2002; 2002US-0411035P.
XX PR 17-SEP-2002; 2002US-0411037P.
XX PR 17-SEP-2002; 2002US-0411041P.
XX PR 17-SEP-2002; 2002US-0411045P.
XX PR 17-SEP-2002; 2002US-0411046P.
XX PR 17-SEP-2002; 2002US-0411048P.
XX PR 17-SEP-2002; 2002US-0411052P.
XX PR 17-SEP-2002; 2002US-0411055P.
XX PR 17-SEP-2002; 2002US-0411073P.
XX PR 17-SEP-2002; 2002US-0411082P.
XX PR 17-SEP-2002; 2002US-0411101P.
XX PR 17-SEP-2002; 2002US-0411111P.
XX PR 18-APR-2003; 2003US-0463700P.
XX PR 18-APR-2003; 2003US-0463708P.
XX PR 18-APR-2003; 2003US-0463716P.
XX PR 18-APR-2003; 2003US-0463732P.
XX PR 02-MAY-2003; 2003US-0467199P.
XX PR 02-MAY-2003; 2003US-0467201P.
XX PR 02-MAY-2003; 2003US-0467203P.
XX PR 02-MAY-2003; 2003US-0467230P.
XX PR 19-MAY-2003; 2003US-0471306P.
XX PR 19-MAY-2003; 2003US-0471336P.
XX PR 22-MAY-2003; 2003US-0472420P.
XX PR 22-MAY-2003; 2003US-0472430P.
XX PR 09-JUN-2003; 2003US-0476609P.

PR 09-JUN-2003; 2003US-0476641P.
PR 08-JUL-2003; 2003US-0485218P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485224P.
PR 14-JUL-2003; 2003US-0485325P.
PR 14-JUL-2003; 2003US-0486446P.
PR 15-JUL-2003; 2003US-0486891P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493341P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
PR 08-AUG-2003; 2003US-0493577P.
XX
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
PI Williams LT, Chu X, Lee E, Hestir K, Beaurang PA, Behrens D;
PI Halenbeck RF, Huang MW, Kothakota S, Haishan L, Linnemann T;
PI Pierce K, Wang Y, Wong JGP, Wu G, Zhang H;
XX
DR WPI; 2004-348438/32.
XX
PT New nucleic acid molecule for diagnosing, preventing or treating diseases
PT such as proliferative (e.g. cancer), inflammatory, immune, metabolic,
PT genetic, bacterial and viral diseases.
XX
PS Claim 1; SEQ ID NO 789; 428pp; English.
XX
CC The present invention relates to an isolated nucleic acid molecule
CC encoding a polypeptide which is believed to be cytostatic,
CC antiinflammatory, immunosuppressive, antibacterial and virucidal. The
CC composition and methods are useful for diagnosing, preventing and
CC treating diseases such as proliferative (e.g. cancer), inflammatory,
CC immune, metabolic, genetic, bacterial and viral diseases. The present
CC sequence represents a human secreted protein encoding sequence. The
CC present sequence is available on WIPOWEB and is not in the specification.
XX
SQ Sequence 558 BP; 202 A; 92 C; 147 G; 117 T; 0 U; 0 Other;

Query Match 50.5%; Score 451.6; DB 12; Length 558;
Best Local Similarity 93.2%; Pred. No. 2.9e-111;
Matches 533; Conservative 0; Mismatches 19; Indels 20; Gaps 5;

QY 87 TGCTCGTGGTTGGAGACCAAGGATAAAGCAAGATCTTATCCACAGACTCCAGGCATA 146
DB 6 TTCTTGTGGTTGGAGACCAAGGATAAAGCAAGATCTTATCCACAGACTCCAGGCATA 65
QY 147 TCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGATGTACTGGAGATGAACACAGA 206
DB 66 TCTTGAAGAACATGCTGAAGAGGAGGCAAAAT--GAAGATGTACTAGAGATGAACAGA 122
QY 207 GGAAGAGAAACAAAGCCCAATTGAGTCCCTGTCAAAGAGAAAGAACCCCTG-AAAAAA 265
DB 123 AGAAGAAAAA-----AACCCCTGTCAAAATGAAGAGAACCCCTGAAAAAAA 169
QY 266 CTGTTGATGTGGCAGCAGAGAAAGTGGTGAATTAATCATCTGAATACCAAGACTG 325
DB 170 CTGTTGATGTGGCAGCAGAGAAAGTGGTGAATTAATCATCTGAATACCAAGACTG 229
QY 326 AGAAGATGCAAGAGAGGCTGAACGATTCAATGTACCTGTGAGCTTGGAGATGAAGAAAG 385
DB 230 AGAAGATGCAAGAGAGGCGCGAACATTCAGTGTACCTGTGAGCTTGGAGATGAAGAAAG 289
QY 386 CTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAACAAAGGTCTCTCATCTGATA 445
DB 290 CTGCTCAGGCAGCTAGGTTGGGATTTCTTCAGTTCCAAACAAAGGTCTCTCATCTG--A 346
QY 446 ACAACCTATGGTTAACTTGGATTAAGCTGAAGGAAGGCTCAAGAGATTTGGTTGAATG 505
DB 347 ACACACCTATGGTTAACTTGGATTAAGCGGAAAGAGCTCAAGAGATTTGGTTGAATG 406
QY 506 TCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAGCGCAT 565
DB 506 TCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAGCGCAT 565

Db 407 TCTCTTCAATTTCCAGAAAAGTCTGAAGATGATAAGAAAAGTGAAGAAAGAGGAGCGCAT 466
QY 566 TTGGGATTTCTCAAGTTTCAGCTTGGAACTGGAAACCAACAGAGGATACAGAGGCAAGAGA 625
Db 467 TTGGGATTTCTCAAGTTTCAGCTTGGAACTGGAAACCAACAGAGGATACAGAGGCAAGAGA 526
QY 626 GGAAAAGAGCAGAGCGCTTTGGGATTTGCCTGA 657
Db 527 GGAAAAGAGCAGAGCGCTTTGGCAATTGCCTGA 558

Search completed: October 4, 2005, 19:50:00
Job time : 581.739 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:19:52 ; Search time 4142.15 Seconds
(without alignments)
10458.088 Million cell updates/sec

Title: US-09-788-476A-1
Perfect score: 894
Sequence: 1 gggagtgagtgagggttaa.....taataaaaaaaaaatagaaaa 894

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	894	100.0	AX405619 Sequence
2	894	100.0	AJ409089 Homo sapi
3	891	99.7	BD276528 FULL-LENG
4	891	99.7	AX048092 Sequence
5	889.4	99.5	BC007099 Homo sapi
6	690.8	77.3	CQ413602 Sequence
7	647.4	72.4	AC146422 Pan trogl
8	633.6	70.9	AC145981 Pan trogl
9	633	70.8	AC145981 Pan trogl
10	622.4	69.6	AC134830 Mus muscu
11	607.6	68.0	AC132088 Mus muscu
12	607.6	68.0	CQ413602 Sequence
13	593.2	66.4	BC027510 Mus muscu
14	593.2	66.4	BC027510 Mus muscu
15	549.6	61.5	AC022200 Homo sapi
16	549.6	61.5	AP001207 Homo sapi
17	549.6	61.5	AC091052 Homo sapi
18	548.6	61.4	AC073328 Homo sapi
19	527	58.9	AR415990 Sequence

AX972824 Sequence
BD111543 EST and e
BV179620 squm1607
AX884873 Sequence
BD024483 Sequence
AC126112 Rattus no
CO684732 Sequence
CO683930 Sequence
CO695331 Sequence
AJ719689 Gallus ga
BX932722 Gallus ga
CO695645 Sequence
CO698322 Sequence
AR427142 Sequence
AX987836 Sequence
BD122695 EST and e
BD122695 EST and e
AX987835 Sequence
AR427141 Sequence
AX987835 Sequence
BD122694 EST and e
CR760849 Xenopus t
AC073487 Homo sapi
AC090678 Homo sapi
AL672298 Mouse DNA
BD071528 Secreted
AR426739 Sequence

AX405619 553 6 AX972824
527 58.9 553 6 BD111543
527 58.9 553 6 BV179620
489.2 52.5 539 11 AX884873
467.6 52.3 471 6 BD024483
467.6 52.3 471 6 AC126112
466.2 52.1 254288 2 CO684732
460.8 51.5 488 6 CO683930
458 51.2 470 6 CO695331
431 48.2 470 6 AJ719689
422 47.2 887 5 BX932722
415.8 46.5 643 5 CO695645
389 43.5 426 6 CO698322
379 42.4 379 6 AR427142
374.2 41.9 405 6 AX987836
374.2 41.9 405 6 BD122695
374.2 41.9 405 6 BD122695
333.6 37.3 383 6 AX987835
333.6 37.3 383 6 BD122694
333.6 37.3 383 6 CR760849
307.2 34.4 797 5 AC073487
281.4 31.5 54666 9 AC090678
275.4 30.8 105736 2 AL672298
263.4 29.5 204050 10 BD071528
258.6 28.9 266 6 AR426739
253.8 28.4 337 6

AX405619 1520 bp DNA linear PAT 14-JUN-2002
Sequence 34 from Patent WO022660.
AX405619
AX405619.1 GI:21438677

ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,
Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.
TITLE Novel nucleic acids and polypeptides
JOURNAL Patent: WO 022660-A 34 21-MAR-2002;
HYSEQ, INC. (US)

FEATURES
source Location/Qualifiers
1..1520
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
636..1268
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD34726.1"
/db_xref="GI:21438678"
/translation="MATETVHLKLAELKOECLARGLETGKIQDLIHLRQAVLEE
HAEENEEDVLGDETEESTKPIELPKVEERPEKTVDAAEKKVKITSEIQT
MQKAEFNPVDSKKAARAFGISSVPTGKSSDNKPMWNLDKLKERARFGLN
VSSISRSKSEDEKLKKRERFIVTSSAGTGTEDTEAKKKRAERFGIA"

ORIGIN
Query Match 100.0%; Score 894; DB 6; Length 1520;
Best Local Similarity 100.0%; Pred. No. 9.5e-203; Mismatches 0; Gaps 0;
Matches 894; Conservative 0; Indels 0; Gaps 0;
Qy 1 GGGAGTGGAGTGGAGGTAACAAGATGCCACCGAGACGGTGGAGCTCCATAAGCTAAAG 60
Db 612 GGGAGTGGAGTGGAGGTAACAAGATGCCACCGAGACGGTGGAGCTCCATAAGCTAAAG 671
Qy 61 CTTCCCGAACTAAAGCAAGAATGTCTTCTGCTGTTGGAGACCAAGGGGAATAAGCAA 120

QY 241 AAAGAGNAGAACCCCTGAAAAAAGCTGTTGATGTGGCAGAGAGAGAAAGTGGTGA 300
 DB 904 AAAGAGAGAGAACCCCTGAAAAAAGCTGTTGATGTGGCAGAGAGAGAAAGTGGTGA 963
 QY 301 ATTACATCTGAAATACACAGACTGAGAGAGATGAGAGAGGCTGAAACGATTCATGTA 360
 DB 964 ATTACATCTGAAATACACAGACTGAGAGAGATGAGAGAGGCTGAAACGATTCATGTA 1023
 QY 361 CCTGTGAGCTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGTTTGGGATTTCTTCAGTT 420
 DB 1024 CCTGTGAGCTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGTTTGGGATTTCTTCAGTT 1083
 QY 421 CCAACAAAGCTCTGTCATCTGATAACAAAGCTATGTTAACTTGGATAAGCTGAAGAA 480
 DB 1084 CCAACAAAGCTCTGTCATCTGATAACAAAGCTATGTTAACTTGGATAAGCTGAAGAA 1143
 QY 481 AGAGCTCAAGATTTGGTTGAAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 540
 DB 1144 AGAGCTCAAGATTTGGTTGAAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAG 1203
 QY 541 AAATGAAAAAGAGGAGGAGCGATTTGGGATTTGTCAAGTTCAGCTGGAACCTGGAACC 600
 DB 1204 AAATGAAAAAGAGGAGGAGCGATTTGGGATTTGTCAAGTTCAGCTGGAACCTGGAACC 1263
 QY 601 ACAGAGATACAGAGGCAAGAGAGGAAAGAGAGAGCGCTTTGGGATTTGCCTGATGA 660
 DB 1264 ACAGAGATACAGAGGCAAGAGAGGAAAGAGAGAGCGCTTTGGGATTTGCCTGATGA 1323
 QY 661 AAAGTCTCGATACCTTTCTGTTCTCCAGTGTTCCTCAATCTCTCTCTCTCTCTCTCT 720
 DB 1324 AAAGTCTCGATACCTTTCTGTTCTCCAGTGTTCCTCAATCTCTCTCTCTCTCTCTCT 1383
 QY 721 ATATATGCTAAATGACAGCTCATGTGCTACGCTCGCTCGCAATGAGGAGCATGTA 780
 DB 1384 ATATATGCTAAATGACAGCTCATGTGCTACGCTCGCTCGCAATGAGGAGCATGTA 1443
 QY 781 CCCAGGTACATCATGAACTGGGCGAGAGTGTGACTTATGCTGTTTCACTTTTAAAG 840
 DB 1444 CCCAGGTACATCATGAACTGGGCGAGAGTGTGACTTATGCTGTTTCACTTTTAAAG 1503
 QY 841 TTGTTGTTGTTTCTGTTTCTGATGATGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTG 894
 DB 1504 TTGTTGTTGTTTCTGTTTCTGATGATGTTGCTGTTGTTGTTGTTGTTGTTGTTGTT 1557

RESULT 3
 LOCUS BD276528 1071 bp DNA linear PAT 17-JUL-2003
 DEFINITION FULL-LENGTH MOLECULES EXPRESSED IN HUMAN TISSUES.
 ACCESSION BD276528
 VERSION BD276528.1 GI:33086296
 KEYWORDS JP 2002543839-A/31.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1071)
 AUTHORS Shih,L., Lu,D.A.M., Lal,P., Batra,S., Tang,T.Y.,
 Yang,J., Azimzai,Y., Reddy,R., Henry, Yue and Baughn,M.R.
 FULL-LENGTH MOLECULES EXPRESSED IN HUMAN TISSUES
 Patent: JP 2002543839-A 31 24-DEC-2002;
 INCYTE GENOMICS INC, Henry YUE, Tom Y TANG, Preeti LAL, Roopa REDDY,
 Sajeev BATRA, Mariah R BAUGHN, Junning YANG, Yalda AZIMZAI, Dyoung Aina
 M LU, Janice AU-YOUNG, Leo L SHIH
 OS Homo sapiens
 PN JP 2002543839-A/31
 PD 24-DEC-2002
 PF 12-MAY-2000 JP 2000618453
 PR 14-MAY-1999 US 09/311937, 14-MAY-1999 US 09/311940, PR
 14-MAY-1999 US 09/311894
 PI leo l shih, janice i au-young, dyoung aina
 m lu, preeti lal, sajeev

PI batra,
 PI tom y tang, junning yang, yalda azimzai, roopa reddy, henry PI
 yue, mariah r baughn
 CC This description about <220> can't be interpreted CC <220>
 CC <221> misc feature
 CC <223> Incyte ID No: 4093555CB1
 FH Key Location/Qualifiers.
 Location/Qualifiers
 1. .1071
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 99.7%; Score 891; DB 6; Length 1071;
 Best Local Similarity 100.0%; Pred. No. 4.9e-202;
 Matches 891; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AGTGGAGTGGGGTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTT 63
 DB 163 AGTGGAGTGGGGTAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTT 222
 QY 64 GCCGAACCTAAAGCAAGAAATGCTTCTGCTGCTGTTGGAGACCAAGGGAATAAAGCAAGAT 123
 DB 223 GCCGAACCTAAAGCAAGAAATGCTTCTGCTGCTGTTGGAGACCAAGGGAATAAAGCAAGAT 282
 QY 124 CTATATCCACAGACTCCAGGCATATCTTTGAAGAAACATGCTGAAGAGAGGCAAAATGAAGAA 183
 DB 283 CTATATCCACAGACTCCAGGCATATCTTTGAAGAAACATGCTGAAGAGAGGCAAAATGAAGAA 342
 QY 184 GATGTACTGGGAGATGAACAGAGGAGAGAGAAACCAAGCCATTTGAGCTCCCTGTCAAA 243
 DB 343 GATGTACTGGGAGATGAACAGAGGAGAGAGAAACCAAGCCATTTGAGCTCCCTGTCAAA 402
 QY 244 GAGGAAGAAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAAT 303
 DB 403 GAGGAAGAAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAAT 462
 QY 304 ACATCTGAAATACACAGACTGAGAGAAATGAGAGAGGCTGAACGATTCATGTTACCT 363
 DB 463 ACATCTGAAATACACAGACTGAGAGAAATGAGAGAGGCTGAACGATTCATGTTACCT 522
 QY 364 GTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGTTTGGGATTTCTTCAGTTCCA 423
 DB 523 GTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGTTTGGGATTTCTTCAGTTCCA 582
 QY 424 ACAAAAGGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGCAAGA 483
 DB 583 ACAAAAGGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGCAAGA 642
 QY 484 GCTCAAAAGATTTGGTGTGAAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAA 543
 DB 643 GCTCAAAAGATTTGGTGTGAAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAA 702
 QY 544 CTGAAAAAGGAGGAGGAGCGATTTGGGATTTGTCACAAAGTTCAGCTGGAACCTGGAACCA 603
 DB 703 CTGAAAAAGGAGGAGGAGCGATTTGGGATTTGTCACAAAGTTCAGCTGGAACCTGGAACCA 762
 QY 604 GAGGATACAGAGGCAAGAGAGGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAGAAA 663
 DB 763 GAGGATACAGAGGCAAGAGAGGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAGAAA 822
 QY 664 GTTCCTGATATCTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTTGTGTCACATA 723
 DB 823 GTTCCTGATATCTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTTGTGTCACATA 882
 QY 724 TATGCTTAAATGACAGTCAATGTCGCTAGCTGCTGCTCGCAATGAGGAGCATGTATCCC 783
 DB 883 TATGCTTAAATGACAGTCAATGTCGCTAGCTGCTGCTCGCAATGAGGAGCATGTATCCC 942
 QY 784 CAGGTACATCCATGAACTCGGCGAGAGTGTGACTTATGCTGTTTTCAGCTTTTAAAGTTG 843
 DB 943 CAGGTACATCCATGAACTCGGCGAGAGTGTGACTTATGCTGTTTTCAGCTTTTAAAGTTG 1002

```
QY      844  TTCTGTTTTTTGTTTTGATTATCTTGCTCTTAATAAAAAAATAGAAAA 894
Db      1003  TTGTGTTTTTGTGTTTTGATTATGTTGCTTGTAAATAAAAAAATAGAAAA 1053

RESULT 4
AX048092      1071 bp      DNA      linear      PAT 15-DEC-2000
LOCUS      Sequence 86 from Patent WO0070047.
DEFINITION      AX048092
ACCESSION      AX048092
VERSION      AX048092.1 GI:11876915
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Yue,H., Tang,Y.T., Lal,P., Reddy,R., Batra,S., Baughn,M.R.,
Yang,J., Azimzai,Y., Lu,D.A., Au-Young,J. and Shih,L.L.
Full-length molecules expressed in human tissues
Patent: WO 0070047-A 86 23-NOV-2000;
Incyte Genomics, Inc. (US)
Location/Qualifiers
1. .1071
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/notes="Incyte ID No: 4093555CB1"

FEATURES
source
Query Match 99.7%; Score 891; DB 6; Length 1071;
Best Local Similarity 100.0%; Pred. No. 4,9e-202;
Matches 891; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4  AGTGGAGTGAGGGGTAACAGATGCGCACCGTGAGCTCCATAAGCTAAAGCTT 63
Db      163  AGTGGAGTGAGGGGTAACAGATGCGCACCGTGAGCTCCATAAGCTAAAGCTT 222
QY      64  GCCGAATAAAGCAAGATGCTTGTCTGTGTTGGAGACCAAGGGAATAAAGCAAGAT 123
Db      223  GCCGAATAAAGCAAGATGCTTGTCTGTGTTGGAGACCAAGGGAATAAAGCAAGAT 282
QY      124  CTTATCCACAGCTCCAGGCATATCTTGAAGACATGCTCAAGAGAGCGCAATGAAGAA 193
Db      283  CTTATCCACAGCTCCAGGCATATCTTGAAGACATGCTCAAGAGAGCGCAATGAAGAA 342
QY      184  GATGTACTGGGAGATGAAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAA 243
Db      343  GATGTACTGGGAGATGAAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAA 402
QY      244  GAGGAAGAACCCCTGTAAGAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAATTT 303
Db      403  GAGGAAGAACCCCTGTAAGAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAATTT 462
QY      304  ACATCTGAATACACAGACTGAGAGATGCGAAGAGGCGCTCAACGATTCAATGTACCT 363
Db      463  ACATCTGAATACACAGACTGAGAGATGCGAAGAGGCGCTCAACGATTCAATGTACCT 522
QY      364  GTGAGCTTGAGAGTAAAGAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCA 423
Db      523  GTGAGCTTGAGAGTAAAGAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCA 582
QY      424  ACAAAGGCTGTCATCTGATTAACAACCTATGTTAACTTGATTAAGCTGAAGGAAAGA 483
Db      583  ACAAAGGCTGTCATCTGATTAACAACCTATGTTAACTTGATTAAGCTGAAGGAAAGA 642
QY      484  GCTCAAGATTGTTGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAA 543
Db      643  GCTCAAGATTGTTGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAA 702
QY      544  CTGAAAAAGAGGAGGAGCGATTGGGATTTGTCAACAGTTTCAAGTTGGAATCGAACCA 603
```

```
Db      703  CTCAAAAAGAGGAAGAGAGCGATTGGGATTGTCAACAAGTTTCAGCTGGAACTGGAACACA 762
QY      604  GAGGATACAGAGCGCAAGAGAGGAAAGAGAGAGCGCTTTGGGATTGCCTGATGAAAA 663
Db      763  GAGGATACAGAGCGCAAGAGAGGAAAGAGAGAGCGCTTTGGGATTGCCTGATGAAAA 822
QY      664  GTTCTCTGATACATTTCTGTTCTCCAGTGTTTTCCATTTCTCTCTTCTTGGTTCACATA 723
Db      823  GTTCTCTGATACATTTCTGTTCTCCAGTGTTTTCCATTTCTCTCTTCTTGGTTCACATA 882
QY      724  TATGCGCTAAATGACAGATCATGTGCTTACGCTGCTGCTGCAATGAGGAGCATGTACCC 783
Db      883  TATGCGCTAAATGACAGATCATGTGCTTACGCTGCTGCTGCAATGAGGAGCATGTACCC 942
QY      784  CAGGTATACATCCATGAACTCGGCGAGCAGTTTGACTTATTGCTGTTTCAGCTTTAAAGGTTG 843
Db      943  CAGGTATACATCCATGAACTCGGCGAGCAGTTTGACTTATTGCTGTTTCAGCTTTAAAGGTTG 1002
QY      844  TTGTGTTTTTTGTTTTGATTATGTTGCTTGTAAATAAAAAAATAGAAAA 894
Db      1003  TTGTGTTTTTTGTTTTGATTATGTTGCTTGTAAATAAAAAAATAGAAAA 1053

RESULT 5
BC007099      923 bp      mRNA      linear      PRI 29-JUN-2004
LOCUS      Homo sapiens cytokine induced protein 29 kDa, mRNA (CDNA clone
DEFINITION      MGC:14726 IMAGE:4273903), complete cds.
ACCESSION      BC007099
VERSION      BC007099.1 GI:13937970
KEYWORDS      MGC.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 923)
Straussberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Altschul,S.F., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Klausner,R.D., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaeitz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carinci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalius,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 923)
Straussberg,R.
Direct Submission
Submitted (30-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
```

Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LENL at: <http://image.llnl.gov>
Series: IRAL Plate: 21 Row: 9 Column: 18
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21361588.

FEATURES

source

Location/Qualifiers

1..923

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:14726 IMAGE:4273903"

/tissue_type="Prostate"

/clone_lib="NIH MGC_83"

/lab_host="DH10B"

/note="Vector: pDNR-LIB"

1..923

/gene="CIP29"

/note="synonyms: MGC14726, HCC-1, HSFC316"

/db_xref="LocusID:84324"

22..654

/gene="CIP29"

/codon_start=1

/product="cytokine induced protein 29 kDa"

/protein_id="AAH07099.1"

/db_xref="GI:13937971"

/db_xref="LocusID:84324"

/translation="MATEVELHKLKLAELKQECLEARGLTKGIKQDLIHLQAYLEE
HAEENEEVLDGETEEETKPIELPEVKEEPEPTVDVAEEKVKVITISEIPQTER
MQRARFNVPVLESKKAAARFGISVPTTKSGDNKPNWNLDKLKERQRFGLN
VSSIRKSEDEKLKKRKRFGIVTSSAGTGTEDTEAKKRKQAEKRGIA"

JOURNAL Biochem. Biophys. Res. Commun. 292 (3), 593-600 (2002)
MEDLINE 21920340
PUBMED 11922608
REFERENCE 2 (bases 1 to 633)
AUTHORS Fukuda, S., Wu, D.W., Stark, K. and Pelus, L.M.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2002) Microbiology/Immunology, Indiana University, School of Medicine, 1044 West Walnut Street, Indianapolis, IN 46202-5121, USA
FEATURES
source
1. .633
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="primary human breast cancer"
CDS
1. .633
/note="nuclear protein with SAP motif"
/codon_start=1
/product="proliferation associated cytokine-inducible protein CIP29"
/protein_id="AA09686.1"
/db_xref="GI:20069116"
/translation="MATETVELHLKLAEKQELARLETKGKQDLIHLQALYEE
HAEEAENEVDLGDTEETETKPIELPKVEEPEKTVDAAEKKVKITSEIPOTER
MOKRAEFNVPVLSLEKKAARAFEGISSVPTKGLSSDNKPMVNLKLERAPFGLN
VSSISRSKSEDEKLKKRERFIVTSAGTGTEDTEAKKRKRERFGIA"
ORIGIN
Query Match 70.8%; Score 633; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 2e-140;
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 25 ATGCGGACCGAGACGGTGGAGCTCATAAGCTAAAGCTTGCCTGCGAAGCTTAAAGCAAGATGT 84
DB 1 ATGCGGACCGAGACGGTGGAGCTCATAAGCTTAAAGCTTGCCTGCGAAGCTTAAAGCAAGATGT 60
QY 85 CTTGCTCGTGTGGTGGAGACCAAGGATTAAGCAAGTCTTATCCACAGACTCCAGCA 144
DB 61 CTTGCTCGTGTGGTGGAGACCAAGGATTAAGCAAGTCTTATCCACAGACTCCAGCA 120
QY 145 TATCTTGAAGAAATGCTGAAGAGGAGCAATGAAGAGATGTACTGGGAGATGAACA 204
DB 121 TATCTTGAAGAAATGCTGAAGAGGAGCAATGAAGAGATGTACTGGGAGATGAACA 180
QY 205 GAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAGGAGGAGAAACCCCTGAGAAA 264
DB 181 GAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAGGAGGAGAAACCCCTGAGAAA 240
QY 265 ACTGTTGATGGCAGCAGAGAGAAAGTGTGAAAATTATCTCTGAAATACCAAGACT 324
DB 241 ACTGTTGATGGCAGCAGAGAGAAAGTGTGAAAATTATCTCTGAAATACCAAGACT 300
QY 325 GAGAGATGCAGAGAGGCTGAACGATTCATCTGCTGAGCTTGGAGGTAAGAAA 384
DB 301 GAGAGATGCAGAGAGGCTGAACGATTCATCTGCTGAGCTTGGAGGTAAGAAA 360
QY 385 GCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGCTGCTCATCTGAT 444
DB 361 GCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGCTGCTCATCTGAT 420
QY 445 AACAACTATATGGTTAATCTTGGATTAAGCTGAAGAAAGAGCTCAAGATTTGGTTTGAAT 504
DB 421 AACAACTATATGGTTAATCTTGGATTAAGCTGAAGAAAGAGCTCAAGATTTGGTTTGAAT 480
QY 505 GTCTCTTCAATCTCCAGAAAGCTGAGATGATGAGAACTGAAAGAGGAGGCGA 564
DB 481 GTCTCTTCAATCTCCAGAAAGCTGAGATGATGAGAACTGAAAGAGGAGGCGA 540
QY 565 TTTGGGATTTGACAAAGTTTCAAGTTCAGTGGAACTGGAACCAAGAGGATACAGAGGCAAGAG 624
DB 541 TTTGGGATTTGACAAAGTTTCAAGTTCAGTGGAACTGGAACCAAGAGGATACAGAGGCAAGAG 600
QY 625 AGGAAAAGACAGAGCGCTTTGGGATTTGCCTGA 657

Db 601 AGGAAAAGACAGAGCGCTTTGGGATTTGCCTGA 633
RESULT 10
AC145981
LOCUS
DEFINITION Pan troglodytes chromosome UNK clone RP43-13G23, *** SEQUENCING IN
PROGRESS ***, 46 unordered pieces.
AC145981 220280 bp DNA linear HTG 01-AUG-2003
AC145981.1 GI:33386932
VERSION HTG; HTGS PHASE1.
KEYWORDS Pan troglodytes (chimpanzee)
SOURCE Pan troglodytes
ORGANISM Pan troglodytes
REFERENCE
1 (bases 1 to 220280)
AUTHORS Wilson, R.K.
TITLE The sequence of Pan troglodytes clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 220280)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: C_PT013G23
* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1338: contig of 1338 bp in length
* 1339: gap of unknown length
* 1438: contig of 1199 bp in length
* 2637: gap of unknown length
* 2737: contig of 1328 bp in length
* 4065: gap of unknown length
* 4165: contig of 1089 bp in length
* 5254: gap of unknown length
* 5354: contig of 1724 bp in length
* 7078: gap of unknown length
* 7178: contig of 1093 bp in length
* 8271: gap of unknown length
* 8372: contig of 1145 bp in length
* 9517: gap of unknown length
* 10945: contig of 1329 bp in length
* 11045: gap of unknown length
* 12864: contig of 1819 bp in length
* 12964: gap of unknown length
* 15017: contig of 2053 bp in length
* 15117: gap of unknown length
* 16693: contig of 1576 bp in length
* 16793: gap of unknown length
* 16794: contig of 1921 bp in length
* 18715: gap of unknown length
* 18815: contig of 2518 bp in length
* 21332: gap of unknown length
* 21433: contig of 1808 bp in length
* 23241: gap of unknown length
* 23341: contig of 1592 bp in length
* 24932: gap of unknown length
* 25032: contig of 2060 bp in length
* 25033: gap of unknown length
* 27092: contig of 2060 bp in length

```
* 27093 27192: gap of unknown length
* 27193 28362: contig of 1170 bp in length
* 28363 28462: gap of unknown length
* 28463 29890: contig of 1428 bp in length
* 29891 31733: contig of 1743 bp in length
* 31734 31833: gap of unknown length
* 31834 34959: contig of 3126 bp in length
* 34960 35059: gap of unknown length
* 35060 37949: contig of 2890 bp in length
* 37950 38049: gap of unknown length
* 38050 41435: contig of 3386 bp in length
* 41436 41535: gap of unknown length
* 41536 46033: contig of 4498 bp in length
* 46034 46134: gap of unknown length
* 46135 49523: contig of 3390 bp in length
* 49524 49623: gap of unknown length
* 49624 54040: contig of 4417 bp in length
* 54041 54140: gap of unknown length
* 54141 58313: contig of 4173 bp in length
* 58314 58413: gap of unknown length
* 58414 62876: contig of 4463 bp in length
* 62877 62976: gap of unknown length
* 62977 69574: contig of 6598 bp in length
* 69575 75666: contig of 5992 bp in length
* 75667 75766: gap of unknown length
* 75767 81362: contig of 5596 bp in length
* 81363 81462: gap of unknown length
* 81463 85272: contig of 3810 bp in length
* 85273 91974: gap of unknown length
* 91975 92074: gap of unknown length
* 92075 96689: contig of 4615 bp in length
* 96690 103977: contig of 7188 bp in length
* 103978 104077: gap of unknown length
* 104078 109657: contig of 5580 bp in length
* 109658 109757: gap of unknown length
* 109758 116796: contig of 6939 bp in length
* 116797 124010: contig of 7214 bp in length
* 124011 124110: gap of unknown length
* 124111 132127: contig of 8017 bp in length
* 132128 132227: gap of unknown length
* 132228 139538: contig of 7311 bp in length
* 139539 139638: gap of unknown length
* 139639 148819: contig of 9181 bp in length
* 148820 148919: gap of unknown length
* 148920 157172: contig of 8253 bp in length
* 157173 157173: gap of unknown length
* 157273 168568: contig of 11296 bp in length
* 168569 168668: gap of unknown length
* 168669 179527: contig of 10859 bp in length
* 179528 179627: gap of unknown length
* 179628 189941: contig of 9314 bp in length
* 189942 189942: gap of unknown length
* 189942 201123: contig of 12081 bp in length
* 201123 201222: gap of unknown length
* 201223 220280: contig of 19058 bp in length.

FEATURES
    source
    1..220280
        /organism="Pan troglodytes"
        /mol_type="genomic DNA"
        /db_xref="taxon:9598"
        /chromosome="UNK"
        /clone="RP43-13G23"
        /size=1338
    misc_feature
    1..1338
        /note="assembly_name:Contig25"
    misc_feature
    1439..2637
        /note="assembly_name:Contig33"
    misc_feature
    2738..4065
        /note="assembly_name:Contig39"
```

```
misc_feature 4166..5254
               /note="assembly_name:Contig40"
misc_feature 5355..7078
               /note="assembly_name:Contig42"
misc_feature 7179..8271
               /note="assembly_name:Contig44"
misc_feature 8372..9516
               /note="assembly_name:Contig45"
misc_feature 9617..10945
               /note="assembly_name:Contig46"
misc_feature 11046..12864
               /note="assembly_name:Contig49"
misc_feature 12965..15017
               /note="assembly_name:Contig51"
misc_feature 15118..16693
               /note="assembly_name:Contig52"
misc_feature 16794..18714
               /note="assembly_name:Contig53"
misc_feature 18815..21332
               /note="assembly_name:Contig54"
misc_feature 21433..23240
               /note="assembly_name:Contig55"
misc_feature 23341..24932
               /note="assembly_name:Contig56"
misc_feature 25033..27092
               /note="assembly_name:Contig57"
misc_feature 27193..28362
               /note="assembly_name:Contig58"
misc_feature 28463..29890
               /note="assembly_name:Contig59"
misc_feature 29991..31733
               /note="assembly_name:Contig60"
misc_feature 31834..34959
               /note="assembly_name:Contig61"
misc_feature 35060..37949
               /note="assembly_name:Contig62"
misc_feature 38050..41435
               /note="assembly_name:Contig63"
misc_feature 41536..46033
               /note="assembly_name:Contig64"
misc_feature 46134..49523
               /note="assembly_name:Contig65"
misc_feature 49624..54040
               /note="assembly_name:Contig66"
misc_feature 54141..58313
               /note="assembly_name:Contig67"
misc_feature 58414..62876
               /note="assembly_name:Contig68"
misc_feature 62977..69574
               /note="assembly_name:Contig69"
misc_feature 69675..75666
               /note="assembly_name:Contig70"
misc_feature 75767..81362
               /note="assembly_name:Contig71"
misc_feature 81463..85272
               /note="assembly_name:Contig72"
misc_feature 85373..91974
               /note="assembly_name:Contig73"
misc_feature 92075..96689
               /note="assembly_name:Contig74"
```

```
Query Match      69.6%; Score 622.4; DB 2; Length 220280;
Best Local Similarity 88.5%; Pred.No. 6.4e-138;
Matches 768; Conservative 0; Mismatches 61; Indels 39; Gaps 7;

QY 12 GAGGGGTAACAGATGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTCCGCAACT 71
      |||||
Db 147627 GAGGGGTAACAGATGCTTAAGCTGAGATGGTGGAGCTCGTAAGCTTCTAACT 147686
      |||||

QY 72 AAAGCAAGATGCTTCTGCTGCTGCTTGGAGACCAAGGGAATAAGCAAGATCTTATCCA 131
      |||||
Db 147687 AAAGCAAGATGCTTCTGCTGCTGCTTGGAGACCAAGGGAATAAGCAAGATCTTATCCA 147746
      |||||
```

```
Qy 132 CAGACTCCAGCATATCTTTGAAGAACATGCTGAGAGGAGGCAATGAAGAGTGACT 191
Db 147747 CAGACTCCAGCATATCTTTGAAGAACATGCTGAGAGGAGGCAAT 147803
Qy 192 GGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGA 251
Db 147804 AGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAAGA 147850
Qy 252 ACCCCCTG-AAAACTGTTGATGTGCGAGCAGAGAAAGTGGTGAATAATACATCTG 310
Db 147851 ACCCCCTGAAAAAACTGTTGATGTGCGAGCAGAGAAAGTGGTGAATAATACATCTG 147910
Qy 311 AATATACCACAGCTGAGAGATGAGAGAGGCGCTGAACGATTCGAATGTTACCTGTGAGCT 370
Db 147911 AATATACCACAGGCTGAGAGATGAGAGAGGCGCGCAACATTCAGTGTACCTGTGAGCT 147970
Qy 371 TGGAGAGTAAG----AAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCCAACA 427
Db 147971 TGGAGAGTAAGATCAAGTAAATCAGGAGCTAGGTTGGGATTTCTTCGAAGCCCAACA 148030
Qy 428 AAGTCTGTCTATGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAAGAGCTC 487
Db 148031 AA-CTCTGTCTATG----AACACACCTATGTTAACTTGGATAAGCCGAAGGAAAGAGCTC 148087
Qy 488 AAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGA 547
Db 148088 AAGATCTGGTTTGAATGCTCTTCAATCTCCGAAAGTCTGAAGATGATAAGAACTGA 148147
Qy 548 AAGAGGGAAGGAGCGGATTTGGGATTTGCAAGTCTCAGCTGGAACCTGGAACCAACAGG 607
Db 148148 AAGAGGGAAGGAGCGGATTTGGGATTTGCAAGTCTCAGCTGGAACCTGGAACCCACAGG 148207
Qy 608 ATACAGAGGCAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTCCTGATGAAAGTTC 667
Db 148208 ATACAGAGGCAAGAGAGGAAAGAGCAGAGCGCTTTGGATTCCTGATGAAAGTTC 148267
Qy 668 CTGATATCTTCTGTTTCCAGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 727
Db 148268 TTGATATCTTCTGTTTCCAGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 148327
Qy 728 CCTAAATGCAGATGTCGCTACGTCCTGCGCAATGAGGAGGAGCATGTACCCAGG 787
Db 148328 CTAAA-----TGCAAGTCTTGCCCTTGCAATGAGGAGGAGCATGTACCCAGG 148371
Qy 788 TACATCATGAATCGCGCAGCAGTTGACTTATGCTTTTTCAGCTTTAAGGTTGTTGT 847
Db 148372 TAAACTGTGAATCGCGCAGCAGTTGACTTATGCTTTTTCAGCTTTAAGGTTGTTGT 148431
Qy 848 GTTTTGTGTTTGAATGTTGCTTGT 875
Db 148432 GTTTTGTGTTTGAATGTTGCTTGT 148459

RESULT 11
AC134830 155666 bp DNA linear ROD 27-NOV-2003
DEFINITION Mus musculus BAC clone RP24-498G16 from chromosome 19, complete
sequence.
ACCESSION AC134830
VERSION AC134830.2 GI:34495140
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
VanBrunt,A.
REFERENCE 1 (bases 1 to 155666)
TITLE The sequence of Mus musculus BAC clone RP24-498G16
AUTHORS Unpublished (2001)
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 155666)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
```

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

3 (bases 1 to 155666)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (01-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 155666)
Wilson,R.K.
Direct Submission
Submitted (07-SEP-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

5 (bases 1 to 155666)
Wilson,R.
Direct Submission
Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 7, 2003 this sequence version replaced gi:23396352.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@wustl.edu
----- Summary Statistics

Center project name: M_BB0498G16

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
<http://genome.wustl.edu>

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

FEATURES

source
1..155666
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="19"
/map="19"
/clone="RP24-498G16"
/clone_lib="RPCI-24"
1..118
/rpt_family="MaLR"
136..410
/rpt_family="MaLR"
472..670
/rpt_family="ERV1"
1983..2092
/rpt_family="MER2_type"
2103..2222
/rpt_family="BC1_MM"
2242..2546
/rpt_family="MER2_type"

repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

Qy	485	CTCAAGAGATTGGTTTTGAAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAC	544
Db	27555	CACAGAGATTGGTTTTGAAATGTCTCTTCCATCTCTAGAAAGTCTGAGGATGATGAGAAC	27614
Qy	545	TGAAAAGAGGAAAGGAGCGATTGGGATTTGTCACAAGTTTCAGCTGGAACTGGAAACACAG	604
Db	27615	TGAAGAAACGGAAGGAGAGATTGGGATTTGTGACAAGTTTCAGCTGGAACTGGAAACACAG	27674
Qy	605	AGGATACAGAGGCAAAAGAGAGGAAAAGACGAGCGCTTTGGGATTGCGCTGATGAAAAAG	664
Db	27675	AGGATACAGAGGCAAAAGAGGAAAAGACGAGCGCTTTGGGAAATTCGATTAATGAAAAG	27734
Qy	665	TTCTGATACATTCTGTCTCTCCAGTGTCTTTTCCATTTCTCTCTCTCTTTGTGTGCATAT	724
Db	27735	TTCTG---CTTTCTGCACCATAGTGTCTTTCCATTTCTCAAAATTTTTCTTGGTTTTAT	27791
Qy	725	ATCGCTAAATGCA CAGTCATGTGCGCTACGTC-----CTGCGCTCGCAATGA	769
Db	27792	ATACATACACACCCCTACCTACGTAATCTATCTACATACACAGTCATGTGCGCTCACAGTGA	27851
Qy	770	GGGAGCATGTACCCAGGTACATCCATGAAGTCTGGCGACAGCTTTGACTTATTGCTGTTTT	829
Db	27852	GGAAGCATGTACGTTTTGGTGACATCCCTTGAATTCAGTAGCAGTTTTATTTTAACTGCTCTTC	27911
Qy	830	CAGCTTTAAGGTTGTTGT--GTTTTCTTTTTTGATTATGTTGCTTGTTAATAAAAAAAT	888
Db	27912	TGCGTTTAGATTGTTGTGTTTTGTTTTTAATTTATTTTGCCTATTAAATAAAAAGGA	27971
Qy	889	AGAAAA	894
Db	27972	AGAAAA	27977

RESULT 12	AC132088	159681 bp	DNA	linear	ROD 27-NOV-2003
LOCUS	Mus musculus BAC clone RP24-410W8	from chromosome 19,	complete		
DEFINITION	sequence.				
ACCESSION	AC132088				
VERSION	AC132088.4	GI:34740419			
KEYWORDS	HTG.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 159681)				
AUTHORS	Harkins,R. and Bielicki,L.				
TITLE	The sequence of Mus musculus BAC clone RP24-410W8				
JOURNAL	Unpublished (2001)				
REFERENCE	2 (bases 1 to 159681)				
AUTHORS	Wilson,R.				
TITLE	Sequencing of Mus musculus				
JOURNAL	Unpublished (2001)				
REFERENCE	3 (bases 1 to 159681)				
AUTHORS	McPherson,J.D. and Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-AUG-2002) Genome Sequencing Center, 4444 Forest Park				
	Parkway, St. Louis, MO 63108, USA				
REFERENCE	4 (bases 1 to 159681)				
AUTHORS	Wilson,R.K.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-JUL-2003) Genome Sequencing Center, 4444 Forest Park				
	Parkway, St. Louis, MO 63108, USA				
REFERENCE	5 (bases 1 to 159681)				
AUTHORS	Wilson,R.K.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-SEP-2003) Genome Sequencing Center, 4444 Forest Park				
	Parkway, St. Louis, MO 63108, USA				
REFERENCE	6 (bases 1 to 159681)				
AUTHORS	Wilson,R.				
TITLE	Direct Submission				
JOURNAL	Submitted (27-NOV-2003) Department of Genetics, Washington				
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				

On Sep 16, 2003 this sequence version replaced gi:331870707

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@watson.wustl.edu

Summary Statistics
Center project name: M BB0410M08

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is overlapped by AC124557 and AC134830.

FEATURES	source	Location/Qualifiers
		1..159681
		/organism="Mus musculus"
		/mol_type="genomic DNA"
		/db_xref="taxon:10090"
		/chromosome="19"
		/map="19"
		/clone="RP24-410M8"
		/clone_lib="RPCI-24"
		1..405
repeat_region		/rpt_family="ERVVK"
		504..738
repeat_region		/rpt_family="MIR"
		1016..1148
repeat_region		/rpt_family="B4"
		1284..1691
repeat_region		/rpt_family="MaLR"
		1696..1839
repeat_region		/rpt_family="MER2_type"
		1912..2118
repeat_region		/rpt_family="L1"
		2160..2247
repeat_region		/rpt_family="MIR"
		2643..2829
repeat_region		/rpt_family="B2"
		2831..3011
repeat_region		/rpt_family="B2"
		complement(2934..3006)
trNA		/product="tRNA-Ser"
		/note="Likely pseudogene"
		SC=6.06)"
repeat_region		4659..4878
		/rpt_family="B2"
repeat_region		5508..5655
		/rpt_family="RMER1B"

```
repeat_region 5743. .5839 /rpt_family="L1"
repeat_region 5924. .6010 /rpt_family="RMER1B"
repeat_region 6496. .6633 /rpt_family="RMER1B"
repeat_region 6779. .6884 /rpt_family="MIR"
repeat_region 6885. .7304 /rpt_family="RMER17C"
repeat_region 7305. .7333 /rpt_family="MIR"
repeat_region 9080. .9263 /rpt_family="L1"
repeat_region 9503. .9905 /rpt_family="MaLR"
repeat_region 12380. .12477 /rpt_family="Alu"
repeat_region 12721. .13257 /rpt_family="MaLR"
repeat_region 14045. .14634 /rpt_family="L1"
repeat_region 16118. .16187 /rpt_family="MER1_type"
repeat_region 16297. .16617 /rpt_family="MaLR"
misc_feature 18720. .19476 /notes="CpG island (GC=72.4, o/e=0.96, #CpGs=99)"
repeat_region 20008. .20080 /rpt_family="B4"
repeat_region 20081. .20114 /rpt_family="Alu"
repeat_region 20300. .20423 /rpt_family="B2"
repeat_region 20495. .21885 /rpt_family="L1"
repeat_region 22810. .22761 /rpt_family="B4"
repeat_region 23159. .23333 /rpt_family="B4"
repeat_region 23761. .23895 /rpt_family="Alu"
repeat_region 25164. .25479 /rpt_family="MaLR"
repeat_region 25575. .25762 /rpt_family="MaLR"
repeat_region 25806. .26115 /rpt_family="L1"
repeat_region 27147. .27426 /rpt_family="B4"
repeat_region 28480. .28600 /rpt_family="MaLR"
repeat_region 28853. .29230 /rpt_family="MaLR"
repeat_region 30803. .30905 /rpt_family="ERV1"
repeat_region 32644. .32851 /rpt_family="MaLR"
repeat_region 33097. .33313 /rpt_family="L1"
repeat_region 33590. .33749 /rpt_family="L1"
repeat_region 33806. .35437 /rpt_family="L1"
repeat_region 35438. .35616 /rpt_family="B2"
repeat_region 35617. .35720 /rpt_family="L1"
repeat_region 35721. .36118 /rpt_family="MaLR"
repeat_region 36119. .36575 /rpt_family="L1"
repeat_region 36823. .37010 /rpt_family="L1"

/rpt_family="B2"
37021. .37166 /rpt_family="B4"
37608. .37699 /rpt_family="ID"
38410. .38547 /rpt_family="Alu"
40185. .40420 /rpt_family="L1"
40579. .40657 /rpt_family="B2"
40693. .40766 /rpt_family="ERV1"
40965. .41155 /rpt_family="MaLR"
41156. .41624

Query Match 68.0%; Score 607.6; DB 10; Length 159681;
Best Local Similarity 82.0%; Pred. No. 2.2e-134;
Matches 743; Conservative 0; Mismatches 144; Indels 19; Gaps 3;

QY 5 GTGGAGTGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTTG 64
Db 30784 GTGGAGTGGGGCGCATCAAGATGGCGCGGCGAGACGGTGGAGCTCCACAAGCTGAAGCTTG 30725
QY 65 CCGNACTAAGCAAGNAATCTTGTCTGCTGTTTGGAGACCAAGGGAATAAAGCAAGATC 124
Db 30724 CTGAACCTAAGCAAGAAATGTCTTGTCTGTTGTTTAGAGACCAAGGGAATAAAGCAAGATC 30665
QY 125 TTATCCACAGACTCCAGGCATATCTTGAAGAACATCTCTGAAGAGGAGGCAAAATGAAGAAG 184
Db 30664 TTATAAATAGGTGCGAGGCTATCTTGAAGACCATCGAGAGAGCAAGCAANTGAAGAAG 30605
QY 185 ATGTACTGGGAGATGAACAGAGGAAGAAAGAAACAAAGCCATTGAGTCCCTGTCAAAG 244
Db 30604 ATGTACTGGGAGATGAACCTGAGGAAGAAAGAAACAAAGCCATAGAACTGCCTGTAAAG 30545
QY 245 AGGAAGAACCCCTGAAAAAATCTGTTGATGTGCGACAGAGAGAAAGTGGTGAATTA 304
Db 30544 AGGAAGAACCTCTGAAAAAGGCTGTTGATGTGCGATCGGAAAAAGGTTGGTAAAGATTA 30485
QY 305 CATCTCAAAATACCACAGACTGAGAGAAATGCAGAAAGGGCTGAACGATTCAATGTACCTG 364
Db 30484 CATCTGGAATACCTCAGACTGAAGAAATGCAGAAAGGGCTGAACGATTCAATGTACCTG 30425
QY 365 TGAGCTTGGAGAGTAAAGAAAGTGTCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAA 424
Db 30424 TAAGCCTGGAGAGTAAAGAAAGGCTGCCCGCGGCTAGGTTGGGAAATTTCTTCAGTTCCAA 30365
QY 425 CAAAAGTCTGTCATCTGATAACAAACCTTATGGTTAACTTGGATAAGCTGAAGGAAAGAG 484
Db 30364 CAAAAGGTTTATCATCTGCACCAAGCCTATGGTTAACTTGGATAAGCTGAAGGAAAGAG 30305
QY 485 CTCAAAAGATTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAAC 544
Db 30304 CACAGAGATTGGTTTGAATGTCTCTTCAATCTCTAGAAAGTCTGAGGATGATGAGAGC 30245
QY 545 TGAAGAAGAGGAAGAGGCGATTGGGATTTGTCACAAGTTTCAGCTGGAACCTGGAACCAAG 604
Db 30244 TGAAGAAGACGGAAGGAGAGATTGGGATTGTGTCAAGTTTCAGCTGGAACCTGGAACCAAG 30185
QY 605 AGGATACAGAGGCAAGAGAGAGAAAGAGCAGAGCGCTTGGGATGGCTGTGTAAGAG 664
Db 30184 AGGATACAGAGGCAAGAGAGAGAAAGAGCAGAGCGCTTGGGATGGCTGTGTAAGAG 30125
QY 665 TTCCTGATATCTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCT 724
Db 30124 TTCGTG---CTTTCTGCACCAATAGTGTCTTCCATTTCTCAATTTCTCAATTTCTTGTGTTTATAT 30068
QY 725 ATGCCTAAATGCACAGTCATGTGCCTACGTC-----CTGCCTCGCAATGA 769
Db 30067 ATACATACACACCCCTACCTACGTAATCTATCTATACATACACAGTCATGTGCCTCACAGTGA 30008
```


Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/INL at: <http://image.llnl.gov>
Series: IRAC Plate: 67 Row: e Column: 6
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13384729.

FEATURES

Location/Qualifiers
1..908
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="MGC:41057 IMAGE:1382954"
/tissue_type="Mammary gland, lactating mouse"
/clone_lib="Soares_mammary_gland_NLMG"
/lab_host="DH10B"
/notes="vector: pT73-Pac"
1..908
/gene="1110005A23Rik"
/product="RIKEN cDNA 1110005A23"
/protein_id="AAH27510.1"
/db_xref="GI:20380578"
/db_xref="LocusID:66118"
/db_xref="MGI:1913368"
5..637
/gene="1110005A23Rik"
/translation="MAEIVELHLKLAELKQECARGLETKGDKDLINRLQAVLED
HAEERANFEDVLGDETEEPKPIELFVEEPEKXADVMSEKKVVKITSGTPQTER
MOKRAERFNVPSLEKKAARAFGISVPTKGLSDTKPMVNLDKLKERAFQGLN
VSSISRSKSEDEKLKKRERFGIVTSSAGTGTTEDEAKKRKRAERFGIA"

ORIGIN

Query Match 66.4%; Score 593.2; DB 10; Length 908;
Best Local Similarity 81.8%; Pred. No. 6.2e-131;
Matches 728; Conservative 0; Mismatches 143; Indels 19; Gaps 3;

Qy	21	CAAGATCGGACCGCAGACGGTGGAGCTCCATAGCTAAAGCTGCCGAACCTAAGCAAGA	80
Db	1	CAAGATCGGCGCGCAGACGGTGGAGCTCCCAAGCTGAAGCTTGCTGAATGAAGCAAGA	60
Qy	81	ATGCTTCTGCTCGTGGTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGACTCCA	140
Db	61	ATGCTTCTGCTCGTGGTTAGAGACCAAGGGAATAAACAAGATCTTATTAATAGGCTGCA	120
Qy	141	GGCATATCTTGAAGAA CATGCTGAAGAGGAGCGCAATGAAGAGATGCTACTGGAGATGA	200
Db	121	GGCCTATCTTGAAGACCATGCAGAAAGAAAGCAAAATGAAGAAGATGCTACTGGAGATGA	180
Qy	201	AACAGAGGAGAGAAACAAAGCCCATTTAGCTCCCTGTCAAGAGGAAGAACCCCTGA	260
Db	181	AACGTGAGGAAGAAACAAAGCCCATAGAACTGCCCTGTTAAAGAGGAAGAACCTCTGA	240
Qy	261	AAAACTGTTGATGTGCAGCAGAGAAAGTGGTGAATAATACATCTGAATACACCA	320
Db	241	AAAGCTGTTGATATGCAATCGGAAGAGAGTGGTGAAGATACATCTGGATACCTCA	300
Qy	321	GACTGAGAGATGCAGAAAGGCGGTGAACGATTCATGTTACCTGTGAGCTTGGAGATGA	380
Db	301	GACTGAAGAATGCAGAAAGGCGGTGAACGATTCATGTTACCTGTGAAGCCTGGAGATGA	360
Qy	381	GAAGCTGCTCGGACAGTGGTGGATTTCTTCAGTTCACAAAGCTCTGTCATC	440
Db	361	GAAGCTGCTCGGCGCGCTAGTGTGGAAATTTCTTCAGTTCACAAAGGTTTATCATC	420
Qy	441	TGATAACAAACCTATGTTAACTTGGATAAGCTCAAGGAAGAGCTCAAGAGATTGGTTT	500
Db	421	TGACACCAAGCCTATGTTAACTTGGATAAGCTCAAGGAAGAGCAAGAGATTGGTTT	480
Qy	501	GAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAATCTGAAGAGGAAGGA	560
Db	481	GAATGCTCTTCCATCTCTAGAAAGTCTGAGGATGATGAGAAGCTGAAGAAACGGAAGGA	540

Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:21:02 ; Search time 181.127 Seconds
(without alignments)
8076.260 Million cell updates/sec

Title: US-09-788-476A-1
Perfect score: 894
Sequence: 1 gggagggactgagggttaa.....taataaaaaaaatagaaaa 894

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	647.4	72.4	22303	4	US-09-949-016-12616, A
C 2	647.4	72.4	22303	4	US-09-949-016-12964, A
C 3	527	58.9	553	4	US-09-621-976-3627, Ap
C 4	467.6	52.3	471	4	US-09-513-999C-736
C 5	374.2	41.9	405	4	US-09-621-976-18639, A
C 6	333.6	37.3	383	4	US-09-621-976-18638, A
C 7	253.8	28.4	337	4	US-09-621-976-18236
C 8	80.6	9.0	7218	1	US-08-232-463-14
C 9	75	8.4	601	4	US-09-949-016-35088
C 10	75	8.4	601	4	US-09-949-016-42456
C 11	74.6	8.3	601	4	US-09-949-016-35089
C 12	74.6	8.3	601	4	US-09-949-016-42457
C 13	53.6	6.0	3211	2	US-08-574-959A-8
C 14	53.6	6.0	3211	3	US-08-357-014-8
C 15	53.6	6.0	3901	2	US-08-574-959A-6
C 16	53.6	6.0	3901	3	US-09-357-014-6
C 17	52.8	5.9	64309	4	US-09-949-016-14581
C 18	52	5.8	92227	4	US-09-949-016-11929
C 19	52	5.8	92232	4	US-09-949-016-15421
C 20	47.6	5.3	929	4	US-09-671-317-14
C 21	46.6	5.2	2754	4	US-09-248-796A-4857
C 22	45.6	5.1	601	4	US-09-949-016-159256
C 23	45.6	5.1	4756	4	US-09-949-016-4455
C 24	45.6	5.1	53737	4	US-09-949-016-16197
C 25	45.4	5.1	7044	4	US-09-949-016-14113
C 26	45.2	5.1	157032	4	US-09-949-016-16502
C 27	45	5.0	601	4	US-09-949-016-90371, A

C 28	45	5.0	601	4	US-09-949-016-90372	Sequence 90372, A
C 29	44.6	5.0	1001	4	US-09-671-317-439	Sequence 439, App
C 30	44.4	5.0	1107	4	US-09-248-796A-4685	Sequence 4685, Ap
C 31	44.2	4.9	6173	4	US-09-949-016-5511	Sequence 5511, Ap
C 32	44.2	4.9	6173	4	US-09-949-016-5512	Sequence 5512, Ap
C 33	44.2	4.9	42672	4	US-09-949-016-17253	Sequence 17253, A
C 34	44.2	4.9	42672	4	US-09-949-016-17253	Sequence 17253, A
C 35	43.8	4.9	18798	4	US-09-949-016-14339	Sequence 14339, A
C 36	43.4	4.9	72549	4	US-09-949-016-16477	Sequence 16477, A
C 37	43.2	4.8	254964	4	US-09-949-016-17392	Sequence 17392, A
C 38	43.2	4.8	254964	4	US-09-949-016-17392	Sequence 17392, A
C 39	43	4.8	2277	1	US-08-676-967-2	Sequence 2, Appli
C 40	43	4.8	2277	1	US-08-676-974-2	Sequence 2, Appli
C 41	43	4.8	2277	2	US-09-098-487-2	Sequence 2, Appli
C 42	43	4.8	4081	3	US-08-999-774A-1	Sequence 1, Appli
C 43	42.6	4.8	5183	1	US-08-459-568-3	Sequence 3, Appli
C 44	42.6	4.8	5183	2	US-08-399-411-3	Sequence 3, Appli
C 45	42.6	4.8	5868	3	US-08-516-859A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-949-016-12616/c
; Sequence 12616, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12616
; LENGTH: 22303
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12616

Query Match	72.4%	Score	647.4	DB	4	Length	22303
Best Local Similarity	90.4%	Pred. No.	6.3e-186	Mismatches	46	Indels	37
Matches	782	Conservative	0				
Qy	12	GAGGGGTAACAAGATGCGACGAGCGGTGGAGCTCCATAAGCTAAAGCTTGCGCAACT	71				
Db	6441	GAGGGGTAACAAGATGCGACGAGCGGTGGAGCTCCATAAGCTAAAGCTTGCGCAACT	6382				
Qy	72	AAAGCAAGATGCTCTTCTCGTCTGTTGGAGACCAAGGGAATAAGCAAGATCTTATCCA	131				
Db	6381	AAAGCAAGATGCTCTTCTCGTCTGTTGGAGACCAAGGGAATAAGCAAGATCTTATCCA	6323				
Qy	132	CAGATCCAGGATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGATGACT	191				
Db	6322	CAGATCCAGGATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGATGACT	6266				
Qy	192	GGGAGATGAACAGAGAGGAAGAACAAAGCCATTTGAGCTCCCTGTGCAAGAGGAGA	251				
Db	6265	AGGAGATGAACAGAGAGGAAGAACAAAGCCATTTGAGCTCCCTGTGCAAGAGGAGA	6219				
Qy	252	ACCCCTTG-AAAAAATGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATACATCTG	310				
Db	6218	ACCCCTTGAAAAAATGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATACATCTG	6159				
Qy	311	AAATACACAGACTGAGAGAAATGCAGAGAGGGCTGAACGATTCATGTACTGTGAGCT	370				

```
Db 6158 AAATACACAGGCTGAGAGATCGAGAGAGGCGCCGAAACAATTCAGTGTACCTGTGAGCT 6099
QY 371 TGGAGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAACAAAAG 430
Db 6098 TGGAGAGTAAGAAAGCTGCTCAGCGAGCTAGGTTTGGGATTTCTTCAGTTTCCAACAAA- 6040
QY 431 GTCTGTCTCATGTATACAAACCTATGGTTAACTTGAAGTGAAGTGAAGGAGCTCAAA 490
Db 6039 GTCTGTCTCATGT--AACACACCTATGGTTAACTTGAAGTGAAGGAGGAGGAGCTCAAA 5982
QY 491 GATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAA 550
Db 5981 GATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAA 5922
QY 551 AGAGGAAGGAGCGATTGGGATTTGTCAAGTTTCAGCTGGAACTGGAACCCACAGAGGATA 610
Db 5921 AGAGGAAGGAGCGATTGGGATTTGTCAAGTTTCAGCTGGAACTGGAACCCACAGAGGATA 5862
QY 611 CAGAGGCAAGAGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTCTCTG 670
Db 5861 CAGAGGCAAGAGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTCTCTG 5802
QY 671 ATACTTTCTCTCTCCAGTCTTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 730
Db 5801 ATACTTTCTCTCTCCAGTCTTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5742
QY 731 AAATGACACAGTCATGTSCCTACGTCCTGCTCGCAATGAGGGAGCATGTACCCAGGTAC 790
Db 5741 AA-----TGCAGTCTTGCCTTGCATGAGGGAGCAGGTACCCAGGTAA 5698
QY 791 ATCCATGAATCGCGCAGCAGTTTGACTTATTCGTTTTCAGCTTTTAAAGTTTGTGTT 850
Db 5697 AACTGTGAATCGCGCGCAGTTTGACTTATTCGTTTTCAGCTTTTAAAGTTTGTGTT 5638
QY 851 TTTGTTTTTGAATGATGTTGTTGTT 875
Db 5637 TTTGTTTTTGAATGATGTTGTTGTT 5613

RESULT 2
US-09-949-016-12964/c
; Sequence 12964, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE OF INVENTION: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 12964
; LENGTH: 22303
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12964

Query Match 72.4%; Score 647.4; DB 4; Length 22303;
Best Local Similarity 90.4%; Pred. No. 6.3e-186;
Matches 782; Conservative 0; Mismatches 46; Indels 37; Gaps 7;

QY 12 GAGGGGTAAACAAGATGCGACCGGAGCGGTGGAGCTCCATAAGCTAAAGCTTCCCGAAT 71
Db 6441 GAGGGGTAAACAAGATGCTGAAGTGGTGGAGCTCCATAAGCTGAAGCTTGTGTAAC 6382
```

```
QY 72 AAAGCAAGAAATCTCTTGTCTGTGGTGGAGACCAAGGGAATAAAGCAAGATCTTATCCA 131
Db 6381 AAAGCAAGAAATCTCTTGTGGTGGAGACCAAGGGAATAAAGCAAGATCTTATCCA 6323
QY 132 CAGACTCCAGGCAATATCTTTGAAGAACATGCTCAAGAGGAGGCAAAATGAAGAGATGTACT 191
Db 6322 CAGACTCCAGGCAATATCTTTGAAGAACATGCTCAAGAGGAGGCAAAAT---GAAGATGTACT 6266
QY 192 GGGAGATGAACAGAGGAGAAAGAAACAAGCCCATTTGAGCTCCCTGTCAAGAGGAGAAGA 251
Db 6265 AGGAGATGAACAGAGGAGAAAGAAACAAGCCCATTTGAGCTCCCTGTCAAGAGGAGAAGA 6219
QY 252 ACCCCCTG-AAAGAACTGTGTGAGTGTGGCAGCAGAGAGAAAGTGGTGAAGATTTACATCTG 310
Db 6218 ACCCCCTGAAAGAACTGTGTGAGTGTGGCAGCAGAGAGAAAGTGGTGAAGATTTACATCTG 6159
QY 311 AAATACCAAGAGCTGAGAGAAATGCAGAGAGGAGGCTGAACGATTTCAATGATGACCTGTGAGCT 370
Db 6158 AAATACCAAGAGCTGAGAGAAATGCAGAGAGGAGGCTGAACGATTTCAATGATGACCTGTGAGCT 6099
QY 371 TGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTTCCAACAAAAG 430
Db 6098 TGGAGAGTAAGAAAGCTGCTCAGGCGAGCTAGGTTTGGGATTTCTTCAGTTTCCAACAAA- 6040
QY 431 GTCTGTCTCATGTATACAAACCTATGGTTAACTTTGGATGAAGCTGAAGGAGAAAGAGCTCAAA 490
Db 6039 GTCTGTCTCATGT--AACACACCTATGGTTAACTTTGGATGAAGGAGGAGAAAGAGCTCAAA 5982
QY 491 GATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAA 550
Db 5981 GATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAA 5922
QY 551 AGAGGAAGAGCGATTTGGGATTTGCAGATTTGCTGGAAGTGAAGGAGGAGGAGGAGGATA 610
Db 5921 AGAGGAAGAGCGATTTGGGATTTGTGCAAGTTTCAGCTGGAACTGGAAACCCACAGAGGATA 5862
QY 611 CAGAGGCAAGAGAGGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTTCTCTG 670
Db 5861 CAGAGGCAAGAGAGGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTTCTCTG 5802
QY 671 ATACTTTCTCTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 730
Db 5801 ATACTTTCTCTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5742
QY 731 AAATGACACAGTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 790
Db 5741 AA-----TGCAGTCTTGCCTTGCATGAGGGAGCAGGTACCCAGGTAA 5698
QY 791 ATCCATGAATCGCGCAGCAGTTTGACTTATTCGTTTTCAGCTTTTAAAGTTTGTGTT 850
Db 5697 AACTGTGAATCGCGCGCAGTTTGACTTATTCGTTTTCAGCTTTTAAAGTTTGTGTT 5638
QY 851 TTTGTTTTTGAATGATGTTGTTGTT 875
Db 5637 TTTGTTTTTGAATGATGTTGTTGTT 5613

RESULT 3
US-09-621-976-3627
; Sequence 3627, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 3627
; LENGTH: 553
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36...452
US-09-621-976-3627

Query Match      58.9%; Score 527; DB 4; Length 553;
Best Local Similarity 99.8%; Pred. No. 3.2e-150;
Matches 538; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy  4 AGTGGAGTGAAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTT 63
Db  15 AGTGGAGTGAAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTT 74
Qy  64 GCCGAACTAAAGCAAGAAATGCTTCTGCTCGTGGTTTGGAGACCAAGGGAATAAACAAGAT 123
Db  75 GCCGAACTAAAGCAAGAAATGCTTCTGCTCGTGGTTTGGAGACCAAGGGAATAAACAAGAT 134
Qy  124 CTTATCCACAGACTCCAGGCAATATCTTGAAGAAATGCTTGAAGAGGAGGCAAAATGAAGAA 183
Db  135 CTTATCCACAGACTCCAGGCAATATCTTGAAGAAATGCTTGAAGAGGAGGCAAAATGAAGAA 194
Qy  184 GATGTAATGAGAGATGAAACAGAGCAAGAAACAAAGCCATTGAGCTCCCTGTCAAA 243
Db  195 GATGTAATGAGAGATGAAACAGAGCAAGAAACAAAGCCATTGAGCTCCCTGTCAAA 254
Qy  244 GAGGAAGAACCCCTGAAAAAATGTTGATGTGCGCAGAGAAAGAAAGTGGTGAAGAAAT 303
Db  255 GAGGAAGAACCCCTGAAAAAATGTTGATGTGCGCAGAGAAAGAAAGTGGTGAAGAAAT 314
Qy  304 ACATCTGAATATACACAGACTGAGAGAAATGCGAAGAGGAGGCTGAACGATTCATGTACCT 363
Db  315 ACATCTGAATATACACAGACTGAGAGAAATGCGAAGAGGAGGCTGAACGATTCATGTACCT 374
Qy  364 GTGAGCTTGGAGAGTAAAGAGCTGCTCGGCACTAGGGTTGGGATTTCTTCAGTTCC 422
Db  375 GTGAGCTTGGAGAGTAAAGAGCTGCTCGGCACTAGGGTTGGGATTTCTTCAGTTCC 434
Qy  423 AACAAAAGTCTGTCTATCTGATAACAAACCTATGTTAACTTGGATAAGCTGGAAGGAAAG 482
Db  435 AACAAAAGTCTGTCTATCTGATAACAAACCTATGTTAACTTGGATAAGCTGGAAGGAAAG 494
Qy  483 AGCTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGA 541
Db  495 AGCTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGA 553

RESULT 4
US-09-513-999C-736
; Sequence 736, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59, US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 736
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25...471
; OTHER INFORMATION:
; US-09-513-999C-736

Query Match      58.9%; Score 527; DB 4; Length 553;
Best Local Similarity 99.8%; Pred. No. 3.2e-150;
Matches 538; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy  4 AGTGGAGTGAAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTT 63
Db  15 AGTGGAGTGAAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTT 74
Qy  64 GCCGAACTAAAGCAAGAAATGCTTCTGCTCGTGGTTTGGAGACCAAGGGAATAAACAAGAT 123
Db  75 GCCGAACTAAAGCAAGAAATGCTTCTGCTCGTGGTTTGGAGACCAAGGGAATAAACAAGAT 134
Qy  124 CTTATCCACAGACTCCAGGCAATATCTTGAAGAAATGCTTGAAGAGGAGGCAAAATGAAGAA 183
Db  135 CTTATCCACAGACTCCAGGCAATATCTTGAAGAAATGCTTGAAGAGGAGGCAAAATGAAGAA 194
Qy  184 GATGTAATGAGAGATGAAACAGAGCAAGAAACAAAGCCATTGAGCTCCCTGTCAAA 243
Db  195 GATGTAATGAGAGATGAAACAGAGCAAGAAACAAAGCCATTGAGCTCCCTGTCAAA 254
Qy  244 GAGGAAGAACCCCTGAAAAAATGTTGATGTGCGCAGAGAAAGAAAGTGGTGAAGAAAT 303
Db  255 GAGGAAGAACCCCTGAAAAAATGTTGATGTGCGCAGAGAAAGAAAGTGGTGAAGAAAT 314
Qy  304 ACATCTGAATATACACAGACTGAGAGAAATGCGAAGAGGAGGCTGAACGATTCATGTACCT 363
Db  315 ACATCTGAATATACACAGACTGAGAGAAATGCGAAGAGGAGGCTGAACGATTCATGTACCT 374
Qy  364 GTGAGCTTGGAGAGTAAAGAGCTGCTCGGCACTAGGGTTGGGATTTCTTCAGTTCC 422
Db  375 GTGAGCTTGGAGAGTAAAGAGCTGCTCGGCACTAGGGTTGGGATTTCTTCAGTTCC 434
Qy  423 AACAAAAGTCTGTCTATCTGATAACAAACCTATGTTAACTTGGATAAGCTGGAAGGAAAG 482
Db  435 AACAAAAGTCTGTCTATCTGATAACAAACCTATGTTAACTTGGATAAGCTGGAAGGAAAG 494
Qy  483 AGCTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGA 541
Db  495 AGCTCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGA 553

RESULT 5
US-09-621-976-18639
; Sequence 18639, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18639
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 126..127
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-18639

Query Match      41.9%; Score 374.2; DB 4; Length 405;
Best Local Similarity 95.0%; Pred. No. 1e-103;
Matches 383; Conservative 11; Mismatches 7; Indels 2; Gaps 1;

Qy  7 GGAGTGAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGGCC 66
Db  1 GGAGTGAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGGCC 60
```

QY 67 GAACTAAAGCAAGATGTC--TTGCTCGTGGTTTGGAGCCAAAGGGAATAAAGCAAGTC 124
Db 61 GAACTAAAGCAAGATGTC--TTGCTCGTGGTTTGGAGCCAAAGGGAATAAAGCAAGTC 120
QY 125 TTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAAG 184
Db 121 TKATVNCAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAAG 180
QY 185 ATGTACTGGAGATGAACAG 244
Db 181 ATGTACTGGAGATGAACAG 240
QY 245 AGGAAGAACCCCTGAAAAAAGCTTGTATGCTGCAGCAGAGAGAGAGAGAGAGAGAGAG 304
Db 241 AGGAAGAACCCCTGAAAAAAGCTTGTATGCTGCAGCAGAGAGAGAGAGAGAGAGAGAG 300
QY 305 CATCTGAAATACACAGACTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 364
Db 301 CATCTGAAATACACAGACTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 365 TGAGCTTGGAGATGAAGAAAGCTGCTGGCGAGCTAGGTTTGG 407
Db 361 TGAGCTTGGAGATGAAGAAAGCTGCTGGCGAGCTAGGTTTGG 403

RESULT 6
US-09-621-976-18638
; Sequence 18638, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18638
; LENGTH: 383
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18638

Query Match 37.3%; Score 333.6; DB 4; Length 383;
Best Local Similarity 99.7%; Pred. No. 2.3e-91;
Matches 333; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 AGTGGAGTGAAGGGTAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTT 63
Db 5 AGTGGAGTGAAGGGTAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAGCTT 64
QY 64 GCCGAACCTAAGCAAGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123
Db 65 GCCGAACCTAAGCAAGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 124
QY 124 CTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAA 183
Db 125 CTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAA 184
QY 184 GATGTACTGGAGATGAACAG 243
Db 185 GATGTACTGGAGATGAACAG 244
QY 244 GAGGAAGAACCCCTGAAAAAAGCTTGTATGCTGCAGCAGAGAGAGAGAGAGAGAGAGAG 303
Db 245 GAGGAAGAACCCCTGAAAAAAGCTTGTATGCTGCAGCAGAGAGAGAGAGAGAGAGAGAG 304
QY 304 ACATCTGAAATACACAGACTGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 337
Db 305 ACATCTGAAATACACAGACTGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAG 338

RESULT 7
US-09-621-976-18236
; Sequence 18236, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18236
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18236

Query Match 28.4%; Score 253.8; DB 4; Length 337;
Best Local Similarity 94.9%; Pred. No. 4.5e-67;
Matches 280; Conservative 3; Mismatches 10; Indels 2; Gaps 2;

QY 601 ACAGAGATACAGAGGCAAG 660
Db 25 ACTTTGGAAACCACTAGCAAG 84
QY 661 AAAGTTCTCTGATACTTCT 720
Db 85 AAAGTTCTCTGATACTTCT 144
QY 721 ATATATGCTCT-AAAATGCACAGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779
Db 145 ATATATGCTCTAAAATGCACAGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 204
QY 780 ACCCAGGTACATCCATGAACCTGCGCAGCAGCAGCTTACCTTATTCGCTGCTTTCAGCTTTAAG 839
Db 205 ACCCAGGTACATCCATGAACCTGCGCAGCAGCAGCTTACCTTATTCGCTGCTTTCAGCTTTAAG 263
QY 840 GTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 894
Db 264 GTTGTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 318

RESULT 8
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; CLONE: 232-463-14
; US-08-232-463-14

Query Match          9.0%; Score 80.6; DB 1; Length 7218;
Best Local Similarity 6.9%; Pred.No.1.9e-13;
Matches 29; Conservative 238; Mismatches 152; Indels 0; Gaps 0;

Qy 1 GGGAGTGAGTGGGGGTAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTAAAG 60
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1439 GGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1380

Qy 61 CTTCCGGAACTAAGCAAGATGTTCTGCTGCTGTTGGAGACCAAGGGAAATGAACAA 120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1379 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1320

Qy 121 GATCTATCCACAGACTCCAGGCATATCTTGAAGACATGCTGAAGAGGCGCAATCAA 180
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1319 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1260

Qy 181 GAAGATGTACTGGGAGTGAACAGAGGAAGAAACAAAGCCCATGAGCTCCCTGTC 240
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1259 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1200

Qy 241 AAAGAGGAAGAACCCCTGAAACAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAA 300
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1199 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1140

Qy 301 ATTACATCTGAATACACAGACTGAGAGATGAGAGAGGCGCTGAACGATTCAATGTA 360
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1139 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1080

Qy 361 CTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTCTTCACT 419
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1079 RRRRRRRRRRRATCGCAGCTCCCTCGACTGCGAGCAAGCTCGGAATTAATCTGT 1021

RESULT 9
US-09-949-016-35088/c
; Sequence 35088, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35088
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-35088

Query Match          8.4%; Score 75; DB 4; Length 601;
Best Local Similarity 8.6%; Pred.No.2.1e-12;
Matches 106; Conservative 0; Mismatches 10; Indels 5; Gaps 2;

Qy 12 GAGGGGTAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCGGA 71
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
415 GAGGGGTAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCGGA 356

Qy 72 AAAGCAA----GAATGCTTCTGCTGCTGTTGGAGACCAAGGGAAATAAGCAAGATCTTA 127
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
355 AAAGTAAGAATGAATGCTTCTGCTGCTGTTGGAGA-CGAGCAATAAAGCAAGATTTT 297

Qy 128 T 128
Db |
296 T 296

RESULT 10
US-09-949-016-42456/c
; Sequence 42456, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 42456
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-42456

Query Match          8.4%; Score 75; DB 4; Length 601;
Best Local Similarity 8.6%; Pred.No.2.1e-12;
Matches 106; Conservative 0; Mismatches 10; Indels 5; Gaps 2;

Qy 12 GAGGGGTAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCGGA 71
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
415 GAGGGGTAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTGCGGA 356

Qy 72 AAAGCAA----GAATGCTTCTGCTGCTGTTGGAGACCAAGGGAAATAAGCAAGATCTTA 127
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
355 AAAGTAAGAATGAATGCTTCTGCTGCTGTTGGAGA-CGAGCAATAAAGCAAGATTTT 297

Qy 128 T 128
Db |
296 T 296

RESULT 11
US-09-949-016-35089/c
; Sequence 35089, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:20:27 ; Search time 3569.42 Seconds
(without alignments)
9533.602 Million cell updates/sec

Title: US-09-788-476A-1
Perfect score: 894
Sequence: 1 999agtggtgaggggttaa.....taataaaaaaaaaatagaaaa 894

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	861	96.3	871	5	BM916484
2	843.6	94.4	874	5	BM915580
3	843.6	94.4	1000	7	CO580484
4	839.8	93.9	877	5	BU849740
5	839	93.8	910	3	AF161434
6	836	93.5	844	3	CR610192
7	831.8	93.0	963	7	CO648068
8	828	92.6	1030	4	BM559381
9	826.6	92.5	895	5	BU174287
10	821	91.8	1070	4	BM460786
11	820.2	91.7	928	5	BU855435
12	817.6	91.5	843	5	BX456776
13	811.4	90.8	817	5	BX456775
14	801.4	89.6	910	5	BU157949
15	800.8	89.6	823	5	BU599301
16	800.4	89.5	914	5	BU508603
17	791.2	88.5	1006	5	BQ068156
18	784.2	87.7	842	3	BQ692014
19	784	87.7	784	3	CR593273
20	783.8	87.7	937	5	BU856660
21	782.2	87.5	883	5	BU146608
22	782	87.5	794	7	CN261593
23	780.8	87.3	795	7	CN261585
24	780.6	87.3	806	4	BG574651

25	774.6	86.6	1105	7	CK231375
26	768.6	86.0	779	5	EX344918
27	765.2	85.6	788	4	BM462456
28	762	85.2	914	5	BU902964
29	759	84.9	783	4	BG910161
30	758.8	84.9	1104	5	BM912983
31	758.4	84.8	876	5	BU855417
c 32	757.2	84.7	785	5	BU632453
33	757	84.7	818	5	BQ227687
34	756.8	84.7	902	5	BQ962276
35	756.6	84.6	783	6	CA775512
36	755.8	84.5	760	4	BM722636
c 37	752	84.1	767	5	BQ008624
38	750.8	84.0	876	5	BU194990
39	749.4	83.8	783	4	BG533012
40	749.2	83.8	791	6	CB956735
c 41	748.2	83.7	753	5	BM979030
42	742.2	83.0	878	5	BQ213428
43	741.4	82.9	887	4	BI255433
44	741	82.9	757	5	BU538304
c 45	740.4	82.8	744	7	CN479544

ALIGNMENTS

RESULT 1
BM916484
LOCUS BM916484 871 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6641858 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5482902
5', mRNA sequence.
ACCESSION BM916484
VERSION BM916484.1 GI:19366863
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 871)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2009 row: 1 column: 07
High quality sequence stop: 748.

FEATURES

source
location/Qualifiers
1..871
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5482902"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 96.3%; Score 861; DB 5; Length 871;

Best Local Similarity 99.3%; Pred. No. 1.1e-208; Matches 864; Conservative 0; Mismatches 6; Indels 0; Gaps 0;	
Qy	25 ATGCGCAGCCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGCCGAATCAAAGCAAGAAATGT 84
Db	1 ATGCGCAGCCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGCCGAATCAAAGCAAGAAATGT 60
Qy	85 CTTGCTCGTGGTTTGGAGACAAGGGAAATAAGCAAGATCTTATCCACAGACTCCAGGCA 144
Db	61 CTTGCTCGTGGTTTGGAGACAAGGGAAATAAGCAAGATCTTATCCACAGACTCCAGGCA 120
Qy	145 TATCTTGAAGAACATGCTGGAAGAGGAGCAATGAAGAAGATGTACTGGAGAGTGAACA 204
Db	121 TATCTTGAAGAACATGCTGGAAGAGGAGCAATGAAGAAGATGTACTGGAGAGTGAACA 180
Qy	205 GAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAGAAACCCCTCGAAAA 264
Db	181 GAGGAAGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGAGGAGAAACCCCTCGAAAA 240
Qy	265 ACTGTTGATGTGGCAGCAGAGAAAGTGTGAAATTAATCATCTGAAATACCAAGACT 324
Db	241 ACTGTTGATGTGGCAGCAGAGAAAGTGTGAAATTAATCATCTGAAATACCAAGACT 300
Qy	325 GAGAGAAATGCAGAAAGAGGGCTGAACGATTCAATGTACTGTGAGCTTGGAGAGTAAGAAA 384
Db	301 GAGAGAAATGCAGAAAGAGGGCTGAACGATTCAATGTACTGTGAGCTTGGAGAGTAAGAAA 360
Qy	385 GCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGTCCTCTCATCTGAT 444
Db	361 GCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGTCCTCTCATCTGAT 420
Qy	445 AACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAGAGCTCAAGAGATTTGGTTGAAAT 504
Db	421 AACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAGAGCTCAAGAGATTTGGTTGAAAT 480
Qy	505 GTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAAGAGCGA 564
Db	481 GTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAAGAGCGA 540
Qy	565 TTTGGGATTTGCACAAGTTACGCTGGAACTGGAAACCAAGAGGATACAGAGGCAAGAAG 624
Db	541 TTTGGGATTTGCACAAGTTACGCTGGAACTGGAAACCAAGAGGATACAGAGGCAAGAAG 600
Qy	625 AGGAAAGAGCAGACGGCTTTGGGATTTGCCCTGATGAAAGTTCCTGATACCTTCTGTCT 684
Db	601 AGGAAAGAGCAGACGGCTTTGGGATTTGCCCTGATGAAAGTTCCTGATACCTTCTGTCT 660
Qy	685 CCAGTGTGTTTCCATTTCT 744
Db	661 CCAGTGTGTTTCCATTTCT 720
Qy	745 GTGCTACGTCCTGCTCGCAATGAGGGAGCATGTACCCAGGTACATCCATGAACCTGG 804
Db	721 GTGCTACGTCCTGCTCGCAATGAGGGAGCATGTACCCAGGTACATCCATGAACCTGG 780
Qy	805 GCAGAGTTTGCATTTATGCTGTTTACGCTTTAAGGTTGTTGTTTTGTTTTGTTTATTA 864
Db	781 GCAGAGTTTGCATTTATGCTGTTTACGCTTTAAGGTTGTTGTTTTGTTTTGTTTATTA 840
Qy	865 TGTGCTGTTGTTAATAAAAAAATAAGAAA 894
Db	841 TGTGCTGTTGTTAATAAAAAAATAAGAAA 870

RESULT 2	BM915580	874 bp	linear	EST 12-MAR-2002
LOCUS	BM915580			
DEFINITION	AGENCOCURT_6701683 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5481778 5', mRNA sequence.			
ACCESSION	BM915580			
VERSION	BM915580.1	GI:19365959		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			


```
RESULT 4
LOCUS BU849740
DEFINITION AGENCOURT_10440969 NIH_MGC_109 Homo sapiens cDNA clone
IMAGE:6598412 5', mRNA sequence.
ACCESSION BU849740
VERSION BU849740.1 GI:24034703
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 877)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M2831 row: c column: 20
High quality sequence stop: 659.
FEATURES
Location/Qualifiers
1..877
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6598412"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 109"
/note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
ORIGIN
Query Match 93.9%; Score 839.8; DB 5; Length 877;
Best Local Similarity 98.6%; Pred. NO. 3e-203;
Matches 855; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
QY 28 GCGACCGAGCGGTGGAGCTCCATAGCTTAAGCTTGCAGCTAAAGCAAGATGCTT 87
DB 1 GCGACCGAGCGGTGGAGCTCCATAGCTTAAGCTTGCAGCTAAAGCAAGATGCTT 60
QY 88 GCTCGTGTGTTGGAGACCAAGGGAATAAGCAAGATCTTATCCAGACTCCAGGCATAT 147
DB 61 GCTCGTGTGTTGGAGACCAAGGGAATAAGCAAGATCTTATCCAGACTCCAGGCATAT 120
QY 148 CTTGAAGAACAATGCTGAAGAGAGGAGCAATGAAGAAGATGTACTGGGAGATGAACAACAGAG 207
DB 121 CTTGAAGAACAATGCTGAAGAGAGGAGCAATGAAGAAGATGTACTGGGAGATGAACAACAGAG 180
QY 208 GAAGAGAAACAAGCCATTGAGCTCCCTCTCAAGAGGAAGACCCCTGAAAAAAT 267
DB 181 GAAGAGAAACAAGCCATTGAGCTCCCTCTCAAGAGGAAGACCCCTGAAAAAAT 240
QY 268 GTTGATGTGGCAGCAGAGAAGAAAGTGTGAAAAATTACATCTGAAATACCAACAGACTGAG 327
DB 241 GTTGATGTGGCAGCAGAGAAGAAAGTGTGAAAAATTACATCTGAAATACCAACAGACTGAG 300
QY 328 AGAATGCAGAAAGGGCTGAACGATTCATGTAATGTACTGTGAGCTTGGAGAGTAAGAAAGCT 387
DB 301 AGAATGCAGAAAGGGCTGAACGATTCATGTAATGTACTGTGAGCTTGGAGAGTAAGAAAGCT 360
```

```
388 GCTCGGCGACCTAGGTTTGGGATTTCTTTCAGTTCCCAACAAAAGGCTCTGTCTCATCTGATAAC 447
361 GCTCGGCGACCTAGGTTTGGGATTTCTTTCAGTTCCCAACAAAAGGCTCTGTCTCATCTGATAAC 420
448 AAACCTATGTTAACTTTGGAATAAGCTGAAGGAAAGAGCTCAAAAGATTTGGTTTCAATGTC 507
421 AAACCTATGTTAACTTTGGAATAAGCTGAAGGAAAGAGCTCAAAAGATTTGGTTTCAATGTC 480
508 TCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAAGAGAGAGGAGCGATTT 567
481 TCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAAGAGAGAGGAGCGATTT 540
568 GGGATGTTCAAAAGTCTCAGCTGGAACTGGAAACCAACAGAGATACAGAGGCAAGAAAGAGG 627
541 GGGATGTTCAAAAGTCTCAGCTGGAACTGGAAACCAACAGAGATACAGAGGCAAGAAAGAGG 600
628 AAAAGAGCAGAGCGCTTTGGGATTTGCCTGATGAAAAGTTCTCTGATATCTTTCTTCTCCA 687
601 AAAAGAGCAGAGCGCTTTGGGATTTGCCTGATGAAAAGTTCTCTGATATCTTTCTTCTCCA 660
688 GTGTTTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 747
661 GTGTTTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
748 CCTACGTCCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATCAATCAATCAATCAATCA 807
721 CCTACGTCCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATCAATCAATCAATCAATCA 780
808 GCAGTTTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 867
781 GCAGTTTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 839
868 TGCTTGTGTAATAAAAAAATAGAAA 894
840 TGCTTGTGTAATAAAAAAATAGAAA 866

RESULT 5
AF161434
LOCUS Homo sapiens HSPC316 mRNA, partial cds.
DEFINITION AF161434
ACCESSION AF161434
VERSION AF161434.1 GI:6841281
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 910)
AUTHORS Ye, M., Zhang, Q.H., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,
Fan, H.Y., Mo, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.
TITLE Human partial CDS from cd34+ stem cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 910)
AUTHORS Ye, M., Zhang, Q.H., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,
Fan, H.Y., Mo, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1999) Shanghai Institute of Hematology, Shanghai
Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,
Shanghai 200025, P. R. China
FEATURES
Location/Qualifiers
1..910
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CBLAIH07"
/cell_type="cd34+ stem cell"
/tissue_type="blood"
<1..769
/codon_start=2
/product="HSPC316"
/protein_id="AAF28994.1"
```

/db_xref="GI:5841282"
/translation="SGVRGNKMATETVELHKLKLAELKOECLARLETKGIKQDLIHR
LQVLEHAEAEANEEDVLGDEEETKPIELPVKEEPEPKTVDAAEKVKVTKTS
EIPQERNKAEARFNVPSLESKKAARAVGISVFNKRSVILITNLMWTLSLKE
ELKDLGLNVSSISRSKDEDEKLKKRERFGIVTSSAGTGTEDTEARRKQESALGLP
DEKFLILSLVQCFPLSFLLTYMPKCTVMCLRPASQ"

ORIGIN

Query Match 93.8%; Score 839; DB 3; Length 910;
Best Local Similarity 99.0%; Pred. No. 4.8e-203;
Matches 886; Conservative 0; Mismatches 5; Indels 4; Gaps 4;

QY 3 GAGTGGAGTGGGGTAAACAGATGCGACCGAGACGGTGGAGCTCCATAGCTAAAGCT 62
DB 1 GAGTGGAGTGGGGTAAACAGATGCGACCGAGACGGTGGAGCTCCATAGCTAAAGCT 60

QY 63 TGCCGAACCTAAAGCAAGAAATGCTTCCTCGTGGTTTGGAGACCAAGGGAATAAGCAAGA 122
DB 61 TGCCGAACCTAAAGCAAGAAATGCTTCCTCGTGGTTTGGAGACCAAGGGAATAAGCAAGA 120

QY 123 TCTTATCCACAGACTCCAGGCATATCTTTGAAGAACATCTGAAGAGGAGGCAAAATGAAGA 182
DB 121 TCTTATCCACAGACTCCAGGCATATCTTTGAAGAACATCTGAAGAGGAGGCAAAATGAAGA 180

QY 193 AGATGTACTGGAGATGAACAGAGGAAGAAGAAACAAAGCCATTGAGCTCCCTGTCAA 242
DB 181 AGATGTACTGGAGATGAACAGAGGAAGAAGAAACAAAGGCCATTGAGCTCCCTGTCAA 240

QY 243 AGAGGAAGAACCCCTGAAAAAATGTTGTATGTGGCAGCAGAGAAAGTGGTGAATAAT 302
DB 241 AGAGGAAGAACCCCTGAAAAAATGTTGTATGTGGCAGCAGAGAAAGTGGTGAATAAT 300

QY 303 TACATCTGAAATACACAGACTGAGAGATGAGAGAGGGCTGAACGATTCATGTACC 362
DB 301 TACATCTGAAATACACAGACTGAGAGATGAGAGAGGGCTGAACGATTCATGTACC 360

QY 363 TGTGAGCTTGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTT-C 421
DB 361 TGTGAGCTTGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCC 420

QY 422 CAACAAAAGGTCTGTGCAT-CTGATAACAAACCTATGTTTAACTTGGATAAGCTGAAGAA 480
DB 421 CAACAAAAGGTCTGTGCATCTCTGATACAAACCTATGTTTAACTTGGATAAGCTTGAAGGA 480

QY 481 AGAGCTCAAGAGATTT-GGTTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGA 539
DB 481 AGAGCTCAAGAGATTTGGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGA 540

QY 540 GAAACTGAAAAAGAGGAAGGAGCGATTTGGGATTTGTCACAGTTTCAGCTGGAACCTGGAAC 599
DB 541 GAAACTGAAAAAGAGGAAGGAGCGATTTGGGATTTGTCACAGTTTCAGCTGGAACCTGGAAC 600

QY 600 CACAGAGATACAGGCAAGAGCAAGAGAAAGAGCAGAGCGCTTGGGATTTGGCTGATG 659
DB 601 CACAGAGATACAGAGGC-AGAAGAGAGAAAGAGCAGAGCGCTTGGGATTTGGCTGATG 659

QY 660 AAAAGTCTCTGATCTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTTCTTCTGTCGA 719
DB 660 AAAAGTCTCTGATCTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTTCTTCTGTCGA 719

QY 720 CATATATGCTAAATGCAAGTCATGTGCTTACGCTTCCTGCTCGCAATGAGGAGCATGT 779
DB 720 CATATATGCTAAATGCAAGTCATGTGCTTACGCTTCCTGCTCGCAATGAGGAGCATGT 779

QY 780 ACCCCAGGTACATCATGACCTGCGGAGCAGTTTGATTTGCTATTTGCTTTTTCAGCTTTAAG 839
DB 780 ACCCCAGGTACATCATGACCTGCGGAGCAGTTTGATTTGCTATTTGCTTTTTCAGCTTTAAG 839

QY 840 GTTGTGTTGTTTTTGTGTTTGTATGTTGTTGTTTAAATAAAAAAATAGAAA 894
DB 840 GTTGTGTTGTTTTTGTGTTTGTATGTTGTTGTTTAAATAAAAAAATAGAAA 894

RESULT 6
CR610192
LOCUS
DEFINITION
full-length cDNA clone CS0CAP004YA11 of Thymus of Homo sapiens (human).
ACCESSION
CR610192
VERSION
CR610192.1 GI:50490999
KEYWORDS
HTC; CNSUT_CDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 844)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS
Li, W.B., Gruber, C., Jesses, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
2 (bases 1 to 844)
Genoscope.
Direct Submission
TITLE
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr
JOURNAL
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES
Location/Qualifiers
source
1..844
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP004YA11"
/tissue_type="Thymus"
/plasmid="pCMVSPORT_6"

ORIGIN

Query Match 93.5%; Score 836; DB 3; Length 844;
Best Local Similarity 100.0%; Pred. No. 2.7e-202; Mismatches 0; Indels 0; Gaps 0;
Matches 836; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 AGCTTGGCGAACTAAAGCAAGAAATGCTTGTCTGTGTTTGGAGACCAAGGGAATAAGC 118
DB 1 AGCTTGGCGAACTAAAGCAAGAAATGCTTGTCTGTGTTTGGAGACCAAGGGAATAAGC 60

QY 119 AAGATCTTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAATG 178
DB 61 AAGATCTTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAATG 120

QY 179 AAGAAGATGTACTGGGAGATGAACAGAGGAAGAAACAAAGCCCATTCAGCTCCCTG 238
DB 121 AAGAAGATGTACTGGGAGATGAACAGAGGAAGAAACAAAGCCCATTCAGCTCCCTG 180

QY 239 TCAAAGAGGAAGAACCCCTCGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGGTGA 298
DB 181 TCAAAGAGGAAGAACCCCTCGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAAGTGGTGA 240

QY 299 AAATTAATCTGAAATACCAAGATGAGAGATGCAAGAGAGGCGTGAACGATTCATG 358
DB 241 AAATTAATCTGAAATACCAAGATGAGAGATGCAAGAGAGGCGTGAACGATTCATG 300

QY 359 TACCTGTGAGCTTGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCAG 418
DB 301 TACCTGTGAGCTTGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTTGGGATTTCTTCAG 360

QY 419 TTCCAAACAAAAGGTCTGCTCATCTGATAACAAACCTATGTTGTTAACTTGCATTAAGCTGAAGG 478
DB 361 TTCCAAACAAAAGGTCTGCTCATCTGATAACAAACCTATGTTGTTAACTTGCATTAAGCTGAAGG 420

QY 479 AAAGAGCTCAAGATTTGGTTTGAATGCTCTCTTCAATCTCCAGAAAGCTCTGAAGATGATG 538

```
|||||
421 AAAGAGCTCAAGATTTGGTTTGAATGTCCTTCAATCTCAGAAAGTCTGAAGATGATG 480
Qy 539 AGAAACTGAAAGAGAGAGGAGCGATTTGGATTGTCACAAAGTTACAGCTGGAACTGGAA 598
Db 481 AGAAACTGAAAGAGAGAGGAGCGATTTGGATTGTCACAAAGTTACAGCTGGAACTGGAA 540
Qy 599 CCACAGAGGATACAGAGCGCAAGAGAGAGAGAGAGAGAGCGCTTTGGGATTCCTTGAT 658
Db 541 CCACAGAGGATACAGAGCGCAAGAGAGAGAGAGAGAGAGCGCTTTGGGATTCCTTGAT 600
Qy 659 GAAAAGTTCCTGATACATTTCTGTTCTCCAGTGTTTCCATTCTCTCTCTCTCTCTCTCT 718
Db 601 GAAAAGTTCCTGATACATTTCTGTTCTCCAGTGTTTCCATTCTCTCTCTCTCTCTCTCT 660
Qy 719 ACATATATGCTAAATGCACAGTCATGTGCTACAGTCTCTGCTCCGCAATGAGGAGCATG 778
Db 661 ACATATATGCTAAATGCACAGTCATGTGCTACAGTCTCTGCTCCGCAATGAGGAGCATG 720
Qy 779 TACCCAGGTACATCCATGACCTGGGAGCAGGAGTTGACCTATTGCTGTTTCAGCTTTAA 838
Db 721 TACCCAGGTACATCCATGACCTGGGAGCAGGAGTTGACCTATTGCTGTTTCAGCTTTAA 780
Qy 839 GGTGTTGCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 894
Db 781 GGTGTTGCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 836
```

```
RESULT 7
CO648068
LOCUS
DEFINITION
ILLUMIGEN MQO 41334 Katze MMPB2 Macaca mulatta cDNA clone
IBUW:25134 5' similar to Bases 527 to 886 highly similar to human
CIP29 (Hs.410597), mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Macaca mulatta (rhesus monkey)
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Macaca.
1 (bases 1 to 963)
Katze,M.G., Thomas,M., Korth,M., Iadonato,S.P. and Magness,C.L.
Large-scale Rhesus Macaque cDNA Sequencing
Unpublished (2003)
Contact: C. Magness
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagness@illumigen.com
Sequenced on 2004.06.02. 750 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCCTCACTAAGGGACAAAAA
BACKWARD: CACTATAGGCGCAATTGGGTA
Insert Length: 963 Std Error: 0.00
Plate: CL000345 row: H column: 08
Seq primer: CCCTCACTAAGGGACAAAAA
POLYA=Yes.
Location/Qualifiers
1..963
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBUW:25134"
/sex="male"
/tissue_type="blood"
/cell_type="PBMC"
```

FEATURES

source

```
/dev_stage="adult"
/lab_host="Electromax DH108"
/clone_lib="Katze MMPB2"
/notes="Vector: pDONR 222; Site 1: BsrG I; Site 2: BsrG I;
Created from CloneMiner cDNA Library Construction kit
(catalog #18249-029)"

ORIGIN

Query Match 93.0%; Score 831.8; DB 7; Length 963;
Best Local Similarity 97.8%; Pred. No. 3.3e-201;
Matches 864; Conservative 0; Mismatches 17; Indels 2; Gaps 2;

Qy 14 GGCGTAACAGATGCGACCGAGACCGTGGAGCTCCATAGCTTAAGCTTGCGGAACTAA 73
Db 3 GGCGTAACAGATGCGACCGAGACCGTGGAGCTCCATAGCTTGCGGAACTAA 62
Qy 74 AGCAAGAAATGTCCTGCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACA 133
Db 63 AGCAAGAAATGTCCTGCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACA 122
Qy 134 GACTCAGGATATCTTGAAGAAATGCTGGAAGAGGAGGCAAAATGAAGAAGATGTACTGG 193
Db 123 GACTCAGGATATCTTGAAGAAATGCTGGAAGAGGAGGCAAAATGAAGAAGATGTACTGG 182
Qy 194 GAGATGAAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 253
Db 183 GAGATGAAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 242
Qy 254 CCCTGAAAGAAATGTTGATGTGGCAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 313
Db 243 CCCTGAAAGAAATGTTGATGTGGCAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 302
Qy 314 TACACAGACTGAGAGAAATGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 373
Db 303 TACCACAGACTGAGAGAAATGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 362
Qy 374 AGAGTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 433
Db 363 AGAGTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 422
Qy 434 TGTCACTCATGATAACAAACCTTATGTTAACTTGGATAAGCTGGAAGGAGGAGGAGGAG 493
Db 423 TGTCACTCATGATAACAAACCTTATGTTAACTTGGATAAGCTGGAAGGAGGAGGAGGAG 482
Qy 494 TTGGTTTGAATGTCCTTCAATCTCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGA 553
Db 483 TTGGTTTGAATGTCCTTCAATCTCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGA 542
Qy 554 GGAAGGAGCGATTTGGGATTTGTCACAAAGTTTCAGCTGGAACCTGGAACCTGGAAC 613
Db 543 GGAAGGAGCGATTTGGGATTTGTCACAAAGTTTCAGCTGGAACCTGGAACCTGGAAC 602
Qy 614 AGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 673
Db 603 AGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 662
Qy 674 CTTTCTGTTCTCCAGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 733
Db 663 CTTATCTGTTCTCCAGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 722
Qy 734 TGCACAGTCAATGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 793
Db 723 TGCACAGTCAATGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 782
Qy 794 CATGAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 852
Db 783 CGTGAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 842
Qy 853 TGTTTTTGATATAT-GTTGCTTGTGTTTAAATAAATAAATAAATAAATAAATAAATAA 894
Db 843 TGTTTTTGATATATGTTGCTGCTGTTTAAATAAATAAATAAATAAATAAATAAATAA 885
```



```
RESULT 8
BM559381
LOCUS
DEFINITION AGENCOURT_6598403 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5474488
5', mRNA sequence.
ACCESSION BM559381
VERSION BM559381.1 GI:18802920
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1030)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTB/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1987 row: m column: 17
High quality sequence stop: 620.
FEATURES
Location/Qualifiers
1..1030
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5474488"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="PH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/Note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
ORIGIN
Query Match 92.6%; Score 828; DB 4; Length 1030;
Best Local Similarity 97.6%; Pred. No. 3.2e-200;
Matches 851; Conservative 0; Mismatches 20; Indels 1; Gaps 1;
QY 22 AAGATGGCGACGACGCTGAGCTCCATAGCTAAAGCTTGCAGCACTAAAGCAAGAA 81
Db 1 AAGATGGCGACGACGCTGAGCTCCATAGCTAAAGCTTGCAGCACTAAAGCAAGAA 60
QY 82 TGTCTTGTCTGCTGTTGGAGACCAAGGAATAAAGCAAGATCTTATCCACAGACTCCAG 141
Db 61 TGTCTTGTCTGCTGTTGGAGACCAAGGAATAAAGCAAGATCTTATCCACAGACTCCAG 120
QY 142 GCATATCTTGAAGACATCTGAGAGGAGGCAATGAAGAAGATGCTATGGAGATGAA 201
Db 121 GCATATCTTGAAGACATCTGAGAGGAGGCAATGAAGAAGATGCTATGGAGATGAA 180
QY 202 ACAGAGGAAGAGAAACAAAGCCATTGAGCTCCCTGTCAAGAGAGGAAGAACCCCTGAA 261
Db 181 ACAGAGGAAGAGAAACAAAGCCATTGAGCTCCCTGTCAAGAGAGGAAGAACCCCTGAA 240
QY 262 AAAAAGCTTTGATGTGGCAGCAGAGAAAGAGTGGTGAAATTTACATCTGAAATACACAG 321
Db 241 AAAAAGCTTTGATGTGGCAGCAGAGAGAAAGTGGTGAAATTTACATCTGAAATACACAG 300
QY 322 ACTGAGAGAAATGAGAGAGGGCTGAACGATTCATGTTACCTGTGAGCTTGGAGATGAG 381
Db 301 ACTGAGAGAAATGAGAGAGGGCTGAACGATTCATGTTACCTGTGAGCTTGGAGATGAG 360
```

```
QY 382 AAAGCTGCTCGGCGACGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAAGGTTCTGTCTCT 441
Db 361 AAAGCTGCTCGGCGACGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAAGGTTCTGTCTCT 420
QY 442 GATAACAACCTATGTTTAACTTCGATAAGCTGAAGGAAAGAGCTCAAGAGATTGCTTTG 501
Db 421 GATAACAACCTATGTTTAACTTCGATAAGCTGAAGGAAAGAGCTCAAGAGATTGCTTTG 480
QY 502 AATGCTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAAGAG 561
Db 481 AATGCTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAAGAG 540
QY 562 CGATTGGGATTTGTCACAAAGTTTCAGCTGGAACCTGGAACCCACAGAGGATACAGAGCAAG 621
Db 541 CGATTGGGATTTGTCACAAAGTTTCAGCTGGAACCTGGAACCCACAGAGGATACAGAGCAAG 600
QY 622 AAGAGGAAAGAGCAGAGCGCTTTGGGATTCGCTGATGAAAGTTCCTGATACCTTTCTGT 681
Db 601 AAGAGGAAAGAGCAGAGCGCTTTGGGATTCGCTGATGAAAGTTCCTGATACCTTTCTGT 660
QY 682 TCTCAGTGTGTTTCCATTTCTCTCCTTCTTCTTGGTCAATATATGCTTAAATGACAGT 741
Db 661 TCTCAGTGTGTTTCCATTTCTCTCCTTCTTCTTGGGACATATATGCTTAAATGACAGT 720
QY 742 CATGTGCTCTACGCTCTGCTCGCAATGAGGAGGAGCATGTACCCAGGATACATCCATGACT 801
Db 721 CAGGTGCTCTACGCTCTGCTCGCAATGAGGAGGAGCATGTACCCAGGATACATCCATGACT 780
QY 802 GCGGCGAGCAGTTGACCTTATTGCTGTTTCAGCTTTAAGGTTGCTGTTGTTTGTGTTTGA 861
Db 781 GCGGCGAGCAGTTGACCTTATTGCTGTTTCAGCTTTAAGGTTGCTGTTGTTTGTGTTTGA 840
QY 862 TT-ATGTTGCTTGTGTTTAAATAAAAAATAGAA 892
Db 841 TTAAGTGTGCTTGTGTTTAAATAAAAAATAGAA 872
RESULT 9
LOCUS
DEFINITION AGENCOURT_7994057 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6082078
5', mRNA sequence.
ACCESSION BUI74287
VERSION BUI74287
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 895)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2310 row: a column: 23
High quality sequence stop: 650.
FEATURES
Location/Qualifiers
1..895
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6082078"
/tissue_type="ductal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
```

```
/clone.lib="NIH_MGC_110"
/notes="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
```

ORIGIN

Query Match	92.5%;	Score 826.6;	DB 5;	Length 895;
Best Local Similarity	96.2%;	Pred. No. 7e-200;		
Matches 844;	Conservative 0;	Mismatches 33;	Indels 0;	Gaps 0;

QY 16 GGTAAACAAGATGGCGACCGACGAGCGGTGGAGCTCCATAAGCTAAAGCTTCCCGAACTAAAG 75

Db 1 GGTAAACAAGATGGCGACCGACGAGCGGTGGAGCTCCATAAGCTAAAGCTTCCCGAACTAAAG 60

QY 76 CAAAGATGCTTCTGCTCGTGGTGGTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGA 135

Db 61 CAAAGATGCTTCTGCTCGTGGTGGTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGA 120

QY 136 CTCAGGCGATATCTTGAAGACATCTGAAGAGGAGGCAAAATGAAGAAGATGTACTGGGA 195

Db 121 CTCAGGCGATATCTTGAAGACATCTGAAGAGGAGGCAAAATGAAGAAGATGTACTGGGA 180

QY 196 GATGAAACAGAGGAAAGAAACAAAGGCCATTTGAGCTCCCTGTCCAAAGAGGAAGAACCC 255

Db 181 GATGAAACAGAGGAAAGAAACAAAGGCCATTTGAGCTCCCTGTCCAAAGAGGAAGAACCC 240

QY 256 CCTGAAACAACTGTTGATGTGCGACAGACAGAAAGAGTGGTGAATTTACATCTGAATA 315

Db 241 CCTGAAACAACTGTTGATGTGCGACAGACAGAAAGTGGTGAATTTACATCTGAATA 300

QY 316 CCACAGACTGAGAGAATCAGAAAGAGGCTGAACCATTTCAATGTACCTGTGAGCTTGGAG 375

Db 301 CCACAGACTGAGAGAATCAGAAAGAGGCTGAACCATTTCAATGTACCTGTGAGCTTGGAG 360

QY 376 AGTAAAGAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTTCCAAACAAAAGGCTG 435

Db 361 AGTAAAGAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTTCCAAACAAAAGGCTG 420

QY 436 TCATCTGATAACAACCTATGTTAACTTGGTAACTGTAAGCTGAAGGAAGAGCTCAAGATTT 495

Db 421 TCATCTGATAACAACCTATGTTAACTTGGTAACTGTAAGCTGAAGGAAGAGCTCAAGATTT 480

QY 496 GGTGTAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAAGAACTGAAAGAGG 555

Db 481 GGTGTAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAAGAACTGAAAGAGG 540

QY 556 AAGAGCGATTTGGGATTTGTCACAGTTTCAGCTGGAATCGGAACACACAGAGGATACAGAG 615

Db 541 AAGAGCGATTTGGGATTTGTCACAGTTTCAGCTGGAATCGGAACACACAGAGGATACAGAG 600

QY 616 GCAAGAGAGGAAAGAGCAGAGCGCTTGGGATTTGGGATTTGGTGAATGAAAGTTCTTGAT 675

Db 601 GCAAGAGAGGAAAGAGCAGAGCGCTTGGGATTTGGGATTTGGTGAATGAAAGTTCTTGAT 660

QY 676 TTCTGTTCTCAGTGTGTTTCCATTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 735

Db 661 TTCTGTTCTCAGTGTGTTTCCATTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 720

QY 736 CACAGTCATGTCCTACGCTCTGCTCGCAATGAGGAGCATGTATACCCAGGATACATCCA 795

Db 721 CACAGTCATGTCCTACGCTCTGCTCGCAATGAGGAGCATGTATACCCAGGATACATCCA 780

QY 796 TGAATCTGGCAGAGTTTGACTTATGCTGTTTTCAGCTTTTAAAGTTGTTGTTTGT 855

Db 781 TGAATCTGGCAGAGTTTGACTTATGCTGTTTTCAGCTTTTAAAGTTGTTGTTTGT 840

QY 856 TTTTGATTTGTTGTTTGAATAAAAAAATAGAA 892

841 GGTTTTGGATTATGNGCGCTTGGTAATANNAAAAA 877

RESULT 10

BM460786

LOCUS

DEFINITION

BM460786 1070 bp mRNA linear EST 05-FEB-2002

AGENCOURT 6421520 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532141

5', mRNA sequence.

ACCESSION

BM460786

VERSION

BM460786.1 GI:18509826

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1070)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12214 row: 0 column: 22

High quality sequence stop: 593.

FEATURES

Location/Qualifiers

1..1070

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5532141"

/tissue_type="leiomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_71"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 2.1 kb. "

ORIGIN

Query Match 91.8%; Score 821; DB 4; Length 1070;

Best Local Similarity 98.2%; Pred. No. 2e-198;

Matches 861; Conservative 0; Mismatches 12; Indels 4; Gaps 3;

QY 15 GGGTAACAAGATGCGACCGACGAGCGGTGGAGCTCCATAAGCTAAAGCTTGCAGAACTAAA 74

Db 4 GGGTAACAAGATGCGACCGACGAGCGGTGGAGCTCCATAAGCTAAAGCTTGCAGAACTAAA 63

QY 75 GCAAGAATGTCTTCTGCTCGTGGTGGAGACCAAGGGGAATAAAGCAAGATCTTATCCACAG 134

Db 64 GCAAGAATGTCTTCTGCTCGTGGTGGAGACCAAGGGGAATAAAGCAAGATCTTATCCACAG 123

QY 135 ACTCCAGCATATCTTGAAGAAATGCTGAAGAGAGGCAAAATGAAGAGATGATCTGGG 194

Db 124 ACTCCAGCATATCTTGAAGAAATGCTGAAGAGAGGCAAAATGAAGAGATGATCTGGG 183

QY 195 AGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAGAGGAGGAACACC 254

Db 184 AGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAGAGGAGGAACACC 243

QY 255 CCCTGAAAAAATCTTGTATGTGGCAGCAGAGAAAGTGGTGAATAATTCATCTGAAAT 314

Db 244 CCCTGAAAAAATCTTGTATGTGGCAGCAGAGAAAGTGGTGAATAATTCATCTGAAAT 303

QY 315 ACCACAGACTGAGAGATGCAAGAGAGGCTGAGAGATTCATGTAATCTGTGAGCTTGA 374

Db 304 ACCACAGACTGAGAGATGCAAGAGAGGCTGAGAGATTCATGTAATCTGTGAGCTTGA 363

QY 375 GAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAAGGCTCT 434


```

RESULT 12
LOCUS      BX456776
DEFINITION BX456776 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004YA11
5-PRIME, mRNA sequence.
ACCESSION  BX456776
VERSION     BX456776.2 GI:47072683
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 843)
AUTHORS   Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On May 22, 2003 this sequence version replaced gi:31036705.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 6407.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0CAP004AA06QPL&c=6407.r.

FEATURES             source
            1..843
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0CAP004YA11"
            /tissue_type="THYMUS"
            /clone_lib="Homo sapiens THYMUS"
            /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
            with a NotI-oligo(dT) primer. Five prime end enriched,
            double-strand cDNA was digested with Not I and cloned into
            the Not I and EcoRV sites of the pCMVSPORT 6 vector.
            Library was not normalized."

ORIGIN
Query Match      91.5%; Score 817.6; DB 5; Length 843;
Best Local Similarity 99.4%; Pred. No. 1.4e-197;
Matches 831; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 59 AGCTTGGCGAATAAGCAAGAAATGTTCTGCTCGTGGTTTGAGACCAGGGAATAAAGC 118
DB 1 AGCTTGGCGAATAAGCAAGAAATGTTCTGCTCGTGGTTTGAGACCAGGGAATAAAGC 60

QY 119 AAGATCTTATCCACAGACTCCAGGCAATCTTTGAAGACATCTCTGAAGAGGCGCAATG 178
DB 61 AAGATCTTATCCACAGACTCCAGGCAATCTTTGAAGACATCTCTGAAGAGGCGCAATG 120

QY 179 AAGAAGATGTTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTG 238
DB 121 AAGAAGATGTTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTG 180

QY 239 TCAAGAGGAAGAACCCCTCGTGAATAAATCTGTTGATGTGGCAGCAGAGAAAGTGGTGA 298
DB 181 TCAAGAGGAAGAACCCCTCGTGAATAAATCTGTTGATGTGGCAGCAGAGAAAGTGGTGA 240

QY 299 AAATTACATCTCAATATACACAGACTGAGAGAATGCAGAGAGGCGTGAACGATTCATG 358
DB 241 AAATTACATCTCAATATACACAGACTGAGAGAATGCAGAGAGGCGTGAACGATTCATG 300

QY 359 TACCTGTGAGCTGGAGAGTAAGAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAG 418
DB 301 TACCTGTGAGCTGGAGAGTAAGAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAG 360

```

```

QY 419 TTCCAACAAAAGGCTCTGTCTCATCTGATAAACAAACCTATGGTTAACTTGGATTAAGCTGAAGG 478
DB 361 TTCCAACAAAAGGCTCTGTCTCATCTGATAAACAAACCTATGGTTAACTTGGATTAAGCTGAAGG 420
QY 479 AAAGAGCTCAAAAGATTTGGTTTGAATGTCCTTCAATCTCCAGAAAGTCTCGAAGATGATG 538
DB 421 AAAGAGCTCAAAAGATTTGGTTTGAATGTCCTTCAATCTCCAGAAAGTCTCGAAGATGATG 480
QY 539 AGAAACTGAAAAGAGAGGAGCGATTGGGATTGTGCACAAAGTTCAGCTGGAACCTGGA 598
DB 481 AGAAACTGAAAAGAGAGGAGCGATTGGGATTGTGCACAAAGTTCAGCTGGAACCTGGA 540
QY 599 CCACAGAGGATACAGAGGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 658
DB 541 CCACAGAGGATACAGAGGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
QY 659 GAAAAGTTCCTGATACATCTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCT 718
DB 601 GAAAAGTTCCTGATACATCTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCT 660
QY 719 ACATATATGCTTAATACAGATGTCATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 778
DB 661 ACATATATGCTTAATACAGATGTCATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
QY 779 TACCCAGGATACATCCATGAACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 838
DB 721 TACCCAGGATACATCCATGAACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 779
QY 839 GGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 894
DB 780 GGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 835

```

```

RESULT 13
LOCUS      BX456775/c
DEFINITION BX456775 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004YA11
3-PRIME, mRNA sequence.
ACCESSION  BX456775
VERSION     BX456775.2 GI:47071640
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
            Full-length cDNA libraries and normalization
            Unpublished (2001)
            On May 22, 2003 this sequence version replaced gi:31034801.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 6407.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0CAP004AA06NPL&c=6407.r.
            Location/Qualifiers
            1..817
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0CAP004YA11"
            /tissue_type="THYMUS"
            /clone_lib="Homo sapiens THYMUS"
            /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
            with a NotI-oligo(dT) primer. Five prime end enriched,
            double-strand cDNA was digested with Not I and cloned into
            the Not I and EcoRV sites of the pCMVSPORT 6 vector.
            Library was not normalized."

FEATURES             source
            1..817

```


Db 421 GCTGAAGAAAGAGAGCTCAAGAGTTGGTTTGAATGCTCTTCAATCTCCAGAAAGCTGA 480
Qy 531 AGATGATGAGAAACTGAAAGAGAGAGGAGCGATTGGGATGTCACAAAGTTTCAGCTGG 590
Db 481 AGATGATGAGAAACTGAAAGAGAGAGGAGCGATTGGGATGTCACAAAGTTTCAGCTGG 540
Qy 591 AACTGGAACCAAGAGGATACAGAGGCAAGAGAGGAAAGAGAGAGCGCTTTGGGAT 650
Db 541 AACTGGAACCAAGAGGATACAGAGGCAAGAGAGGAAAGAGAGAGCGCTTTGGGAT 600
Qy 651 TGCCTGATGAAAGTTCTCTGATACATCTCTGTTCTCCAGTGTCTTCCATTTCTCTCTTCT 710
Db 601 TGCCTGATGAAAGTTCTCTGATACATCTCTGTTCTCCAGTGTCTTCCATTTCTCTCTTCT 660
Qy 711 TCTTGGTCAATATATGCTTAATGCAAGTATGCTGCTCGCTCGCTCGCAATGAG 770
Db 661 TCTTGGTCAATATATGCTTAATGCAAGTATGCTGCTCGCTCGCTCGCAATGAG 720
Qy 771 GGAGCATGATCCAGGATACATCCATGAATGCTGCGGACAGTTTGACTTATGCTGTTTC 830
Db 721 GGAGCATGATCCAGGATACATCCATGAATGCTGCGGACAGTTTGACTTATGCTGTTTC 780
Qy 831 AGCTTTAAGGTG-TTGTGTTTCTGTTTGTGATTGTTGTTGTTAATAAAAAAAT 888
Db 781 AGCTTTAAGGTGTTTGGGTTTGGTTTGGATAGTGGCTGGTACCAACAGATT 839

RESULT 15
BU599301
LOCUS AGENCOURT_8908959 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6458323
DEFINITION 5', mRNA sequence.
ACCESSION BU599301
VERSION BU599301.1 GI:23251060
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 823)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCW2641 row: j column: 20
High quality sequence stop: 602.
Location/Qualifiers
1. 823

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6458323"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_142"
/note="Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccctcgcc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:
5'-RAGCAGTGGTATCAACGAGAGTGGCCATTAGCGCGGG-3' and

5'-ATTCTAGAGCCGAGCGCGGCATG-dt(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the >0.5 kb size fraction (other fractions present in NIH_MGC_141). Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 89.6%; Score 800.8; DB 5; Length 823;
Best Local Similarity 99.5%; Pred. No. 2.6e-193;
Matches 813; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
Qy 4 AGTGGAGTGGGGTAAACAAGATGGCCAGCAGAGCGGTGGAGCTCCATAAGCTAAAGCTT 63
Db 6 AGTGGAGTGGGGGTAACAAGATGGCCAGCAGCGGTGGAGCTCCATAAGCTAAAGCTT 65
Qy 64 GCCGAACTAAAGCAAGAAATGTCTTGTCTGTGGTGGAGACCAAGGGAATAAACAAGAT 123
Db 66 GCCGAACTAAAGCAAGAAATGTCTTGTCTGTGGTGGAGACCAAGGGAATAAACAAGAT 125
Qy 124 CTATCCACAGACTCCAGGCAATATCTTTGAAGAACATCTGGAAGAGGAGGCAAAATGAAGAA 183
Db 126 CTATCCACAGACTCCAGGCAATATCTTTGAAGAACATCTGGAAGAGGAGGCAAAATGAAGAA 185
Qy 184 GATGTACTGGAGATGAACAAGAGGAAGAACAAGCCCATTTGAGCTCCCTGTCAA 243
Db 186 GATGTACTGGAGATGAACAAGAGGAAGAACAAGCCCATTTGAGCTCCCTGTCAA 245
Qy 244 GAGGAAGAAACCCCTGTAAGAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAAGATT 303
Db 246 GAGGAAGAAACCCCTGTAAGAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAAGATT 305
Qy 304 ACATCTCAAAATACACAGACTGAGAGAAATGCAGAAAGGCGCTGAACCAATGTACCT 363
Db 306 ACATCTCAAAATACACAGACTGAGAGAAATGCAGAAAGGCGCTGAACCAATGTACCT 365
Qy 364 GTGAGCTTGGAGATGAAGAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCA 423
Db 366 GTGAGCTTGGAGATGAAGAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCA 425
Qy 424 ACAAAGGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAGA 483
Db 426 ACAAAGGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAGA 485
Qy 484 GCTCAAAAGATTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAA 543
Db 486 GCTCAAAAGATTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAA 545
Qy 544 CTGAAAAGAGGAGGAGCGATTGGGATTTGTCAACAGTTTCAGCTGGAACCTGGAACCCACA 603
Db 546 CTGAAAAGAGGAGGAGCGATTGGGATTTGTCAACAGTTTCAGCTGGAACCTGGAACCCACA 605
Qy 604 GAGGATACAGAGGCAAAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCCGTGATGAAA 663
Db 606 GAGGATACAGAGGCAAAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCCGTGATGAAA 665
Qy 664 GTTCCTGATACCTTTCTGTTCTCAGTGTGTTTCCATTTCTCTCTCTTCTTCTTGGTCACATA 723
Db 666 GTTCCTGATACCTTTCTGTTCTCAGTGTGTTTCCATTTCTCTCTCTTCTTCTTGGTCACATA 725
Qy 724 TATGCCCTAAATGACAGCTGATGTCCTACGTCCTGCTCGCAATGAGGAGCATGTACCC 783
Db 726 TATGCCCTAAATGACAGCTGATGTCCTACGTCCTGCTCGCAATGAGGAGCATGTACCC 785
Qy 784 CAGGTACATCCAT-GAACTGCGGACAGCAGTTTGACTT 819
Db 786 CAGGTACATCCATGGAACCTGCGGACAGCAGTTTGACTT 822

Search completed: October 5, 2005, 00:04:19
Job time : 3579.42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 00:04:33 ; Search time 97 Seconds
(without alignments)
837.316 Million cell updates/sec

Title: US-09-788-476a-2

Perfect score: 1040

Sequence: 1 MATETVELHLKLAELKQEC.....GTTEDTEAKRKRAERFGIA 210

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	1040	100.0	210	4 AAB36609	Human FLE
2	1040	100.0	210	5 ABB97210	Abg72881 Novel hum
3	1040	100.0	210	5 ABG72881	Abg72881 Novel hum
4	1040	100.0	228	4 AAU18233	Abg18233 Novel hum
5	1040	100.0	228	5 ABG92654	Abg92654 Human DNA
6	1040	100.0	228	7 ADC25371	ADC25371 Human ext
7	746	71.7	185	8 ADF30022	ADP30022 Human sec
8	742	71.3	149	3 AAG00732	ADP00732 Human sec
9	222	21.3	308	4 ABB63720	Abb63720 Drosophil
10	155	14.9	734	7 ADE14358	Adel14358 Human int
11	155	14.9	747	8 ADS10484	Adsl10484 Human the
12	155	14.9	795	4 ABB61567	Abb61567 Drosophil
13	139	13.4	185	3 AAG20995	Agg20995 Arabidops
14	139	13.4	214	3 AAG20700	Agg20700 Arabidops
15	138	13.3	750	8 ADRO8732	Adr08732 Human pro
16	138	13.3	1150	4 AAM40294	Aam40294 Human pol
17	137.5	13.2	540	4 AAG74992	Agg74992 Human col
18	137.5	13.2	1150	5 ABG32465	Abg32465 Human pro
19	137.5	13.2	1150	8 ADS88261	Ads88261 Human pro
20	136.5	13.1	718	4 AAB92973	Abg92973 Human pro
21	136.5	13.1	718	8 ADS88419	Adg88419 Human pro
22	136.5	13.1	1150	7 ADF50152	Adf50152 Human dea
23	133.5	12.8	633	3 AAG31342	Agg31342 Arabidops
24	133.5	12.8	633	8 ADN72455	Adn72455 Thale cre
25	133	12.8	1146	7 ADF50150	Adf50150 Murine de

ALIGNMENTS

RESULT 1

AAB36609

ID AAB36609 standard; protein; 210 AA.

XX AAB36609;

XX AC AAB36609;

XX DT 09-MAR-2001 (first entry)

XX DE Human FLEXHT-31 protein sequence SEQ ID NO:31.

XX DE Human FLEXHT-31 protein sequence SEQ ID NO:31.

XX KW Human; FLEXHT; full-length molecules expressed in human tissue;
XX KW diagnosis; gene expression; genetic linkage; genetic variability;
XX KW antianemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;
XX KW cytotatic; antiasthmatic; antiinflammatory; hepatotropic; antidiabetic;
XX KW anti-gout; antithyroid; neuroprotective; antiarthritic; osteopathic;
XX KW antipsoriatic; antirheumatic; antiulcer; gene therapy; anaemia; gout;
XX KW epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;
XX KW cancer; immunological disorder; asthma; bronchitis; cirrhosis;
XX KW Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;
XX KW osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;
XX KW ulcerative colitis.

XX OS Homo sapiens.

XX PN WO200070047-A2.

XX PD 23-NOV-2000.

XX PF 12-MAY-2000; 2000WO-US013299.

XX PR 14-MAY-1999; 99US-00311894.

XX PR 14-MAY-1999; 99US-00311937.

XX PR 14-MAY-1999; 99US-00311940.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PA Yue H, Tang YT, Lal P, Reddy R, Batra S, Baughn MR, Yang J;

XX PI Azimzai Y, Lu DAM, Au-Young J, Shih LL;

XX XX WPI; 2001-016234/02.

XX DR N-PSDB; AAC88100.

XX DR Human FLEXHT protein and DNA sequences, useful for treating immunological

XX PT disorders, developmental disorders, and cancers.

XX XX Claim 1; Page 120; 168pp; English.

XX PS AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules

XX CC

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC expressed in human tissues) proteins given in AAB36579 to AAB36633. The
CC present invention describes an isolated polypeptide (A) comprising an
CC amino acid sequence selected from one of 55 amino acid sequences 42-876
CC residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %
CC identical sequence, and a biologically active or immunogenic fragment of
CC the sequence. The FLEXHT proteins can have antianaemic, anticonvulsant,
CC antiarteriosclerotic, immunomodulatory, cytostatic, antiaesthmatic,
CC antiinflammatory, hepatotropic, antidiabetic, anti-gout, antithyroid,
CC neuroprotective, antiarthritic, osteopathic, antipsoriatic, antitumor and
CC antirheumatic activities, and can be used in gene therapy. The
CC polynucleotide sequences can be used to express the protein sequences.
CC Pharmaceutical compositions comprising FLEXHT can be used to treat
CC diseases or conditions associated with altered expression of functional
CC FLEXHT. The proteins and polynucleotides can be used to diagnose and
CC treat disorders including anaemia, epilepsy, arteriosclerosis, and
CC atherosclerosis, developmental disorders, cancers, and immunological
CC disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,
CC diabetes mellitus, gout, Grave's disease, multiple sclerosis,
CC osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and
CC ulcerative colitis
XX
SQ Sequence 210 AA;

Query Match 100.0%; Score 1040; DB 4; Length 210;
Best Local Similarity 100.0%; Pred. No. 1.5e-85;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAEANEEDVLGDET 60
DB 1 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAEANEEDVLGDET 60
QY 61 EEEETKPIELPVKEEPEPTVDVAAEKVKVITSEIPQTERMQKRAERFNVPSLESKK 120
DB 61 EEEETKPIELPVKEEPEPTVDVAAEKVKVITSEIPQTERMQKRAERFNVPSLESKK 120
QY 121 AAAAAFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
DB 121 AAAAAFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
QY 181 FGIVTSAGTGTTEDTEAKKRAERFGIA 210
DB 181 FGIVTSAGTGTTEDTEAKKRAERFGIA 210

RESULT 2
AB97210
ID ABB97210 standard; protein; 210 AA.
XX
AC ABB97210;
XX
DT 28-JUN-2002 (first entry)
XX
DE Novel human protein SEQ ID NO: 478.
XX
DE Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.
XX
OS Homo sapiens.
XX
FN W0200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US026015.
XX
PR 11-SEP-2000; 2000US-00659671.
XX
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI: 2002-292408/33.
DR N-PSDB; ABN32396.
XX
PT An isolated polynucleotide for treating diseases associated with its
encoded polypeptide such as cancer and multiple sclerosis.
XX
PS Example 2; SEQ ID NO 478; 509pp; English.
XX
CC The present invention provides the protein and coding sequences of 444
novel human proteins. These were isolated from expressed sequences tags
(ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention
XX
SQ Sequence 210 AA;

Query Match 100.0%; Score 1040; DB 5; Length 210;
Best Local Similarity 100.0%; Pred. No. 1.5e-85;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAEANEEDVLGDET 60
DB 1 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAEANEEDVLGDET 60
QY 61 EEEETKPIELPVKEEPEPTVDVAAEKVKVITSEIPQTERMQKRAERFNVPSLESKK 120
DB 61 EEEETKPIELPVKEEPEPTVDVAAEKVKVITSEIPQTERMQKRAERFNVPSLESKK 120
QY 121 AAAAAFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
DB 121 AAAAAFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
QY 181 FGIVTSAGTGTTEDTEAKKRAERFGIA 210
DB 181 FGIVTSAGTGTTEDTEAKKRAERFGIA 210

RESULT 3
ABG72881
ID ABG72881 standard; protein; 210 AA.
XX
AC ABG72881;
XX
DT 03-MAR-2003 (first entry)
XX
DE Novel human protein HCC-1.
XX
KW Human; hcc-1; human hepatocellular carcinoma tissue; HCC-M cell line;
KW cancer; hepatocellular carcinoma; antisense gene therapy.
XX
OS Homo sapiens.
XX
PN US2002107190-A1.
XX
PD 08-AUG-2002.
XX
PF 21-FEB-2001; 2001US-00788476.
XX
PR 25-FEB-2000; 2000US-0185116P.
XX
PA (CHUN/) CHUNG C M.
PA (CHAN/) CHAN L.
PA (OUKK/) OU K.
PA (ONGS/) ONG S.
PA (SEOW/) SEOW T K.
PA (LIAN/) LIANG C R.
PA (CHOO/) CHOONG M L.


```

PA (TANL/) TAN L K.
XX
PI Chung CM, Chan L, Ou K, Ong S, Seow TK, Liang CR, Choong ML;
PI Tan LK;
XX
PI WPI; 2002-697878/75.
XX
DR N-PSDB; ABX13935.
XX
XX New nucleic acid which is differentially expressed in human
PT hepatocellular carcinoma tissue useful for diagnosing and developing
PT therapy for hepatocellular carcinoma and related conditions.
XX
XX Claim 4; Fig 2; 23pp; English.
XX
XX The invention describes an isolated nucleic acid whose expression is
CC differential or preferential in human hepatocellular carcinoma tissue or
CC tissue from a related cancer relative to other tissue in the subject(s)
CC diagnosed with the condition. The nucleic acid is used to diagnose and
CC treat hepatocellular carcinoma and related cancers, or modulate one or
CC more activities in a cell e.g. by antisense gene therapy. This is the
CC amino acid sequence of the novel human protein HCC-1 identified from the
CC HCC-M cell and proposed to be involved in nucleic acid binding and
CC transcription control
XX
XX Sequence 210 AA;
SQ
Query Match 100.0%; Score 1040; DB 5; Length 210;
Best Local Similarity 100.0%; Pred. No. 1.5e-85;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MATETVHLKLAELKOECLARGLETGKTKQDILHRLQAYLEHAEANEEDVLGDET 60
Db 1 MATETVHLKLAELKOECLARGLETGKTKQDILHRLQAYLEHAEANEEDVLGDET 60
Qy 61 EEETKTEIPLVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERFNPVLSLSKK 120
Db 61 EEETKTEIPLVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERFNPVLSLSKK 120
Qy 121 AARAARFGISSVPTKGLSSDNKPMVNDLKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
Db 121 AARAARFGISSVPTKGLSSDNKPMVNDLKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
Qy 181 FGIVTSSAGTGTEDTEAKKRAERFGIA 210
Db 181 FGIVTSSAGTGTEDTEAKKRAERFGIA 210
RESULT 4
AAU18233
ID AAU18233 standard; protein; 228 AA.
XX
AC AAU18233;
XX
DT 21-NOV-2001 (first entry)
XX
DE Novel human DNA-binding protein #80.
XX
XX Human; DNA-binding protein; histone; chromo domain protein;
XX chromatin organisation modifier; Y-box binding protein; DNA organisation;
XX gene transcription; malignant disease; autoimmune disorder;
XX rheumatic disease; genetic abnormality; infectious disease;
XX neurological disorder; gene therapy; immunomodulatory; anti-HIV;
XX anti rheumatic; anti microbial; cytostatic.
XX
OS Homo sapiens.
XX
XX WO200155162-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001305.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR

```

PR	02-OCT-2000;	2000US-0237038P.	XX	Claim 11; SEQ ID NO 218; 561pp; English.
PR	02-OCT-2000;	2000US-0237039P.	PS	The present invention relates to the isolation of novel DNA-binding proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding for these proteins. DNA-binding proteins such as histones, chromo (chromatin organisation modifier) domain proteins, and Y-box binding proteins may contribute to diseases resulting from aberrant DNA organisation and/or gene transcription. The sequences of the invention are useful in screening assays to identify antagonists and/or agonists that may enhance or block activities mediated by DNA-binding proteins. Blockers of DNA-binding proteins may be useful in treating disorders such as malignant diseases (e.g. cancer), autoimmune disorders (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's disease). The polynucleotide sequences of the invention may also be used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding proteins. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
PR	02-OCT-2000;	2000US-0237040P.	XX	
PR	13-OCT-2000;	2000US-0239935P.	CC	
PR	13-OCT-2000;	2000US-0239937P.	CC	
PR	20-OCT-2000;	2000US-0240960P.	CC	
PR	20-OCT-2000;	2000US-0241221P.	CC	
PR	20-OCT-2000;	2000US-0241785P.	CC	
PR	20-OCT-2000;	2000US-0241786P.	CC	
PR	20-OCT-2000;	2000US-0241787P.	CC	
PR	20-OCT-2000;	2000US-0241808P.	CC	
PR	20-OCT-2000;	2000US-0241809P.	CC	
PR	20-OCT-2000;	2000US-0241826P.	CC	
PR	01-NOV-2000;	2000US-0244617P.	CC	
PR	08-NOV-2000;	2000US-0246474P.	CC	
PR	08-NOV-2000;	2000US-0246475P.	CC	
PR	08-NOV-2000;	2000US-0246527P.	CC	
PR	08-NOV-2000;	2000US-0246528P.	CC	
PR	08-NOV-2000;	2000US-0246532P.	CC	
PR	08-NOV-2000;	2000US-0246609P.	CC	
PR	08-NOV-2000;	2000US-0246610P.	CC	
PR	08-NOV-2000;	2000US-0246611P.	CC	
PR	08-NOV-2000;	2000US-0246613P.	CC	
PR	17-NOV-2000;	2000US-0249207P.	CC	
PR	17-NOV-2000;	2000US-0249208P.	CC	
PR	17-NOV-2000;	2000US-0249209P.	CC	
PR	17-NOV-2000;	2000US-0249210P.	CC	
PR	17-NOV-2000;	2000US-0249211P.	CC	
PR	17-NOV-2000;	2000US-0249212P.	CC	
PR	17-NOV-2000;	2000US-0249213P.	CC	
PR	17-NOV-2000;	2000US-0249214P.	CC	
PR	17-NOV-2000;	2000US-0249215P.	CC	
PR	17-NOV-2000;	2000US-0249216P.	CC	
PR	17-NOV-2000;	2000US-0249217P.	CC	
PR	17-NOV-2000;	2000US-0249218P.	CC	
PR	17-NOV-2000;	2000US-0249244P.	CC	
PR	17-NOV-2000;	2000US-0249245P.	CC	
PR	17-NOV-2000;	2000US-0249264P.	CC	
PR	17-NOV-2000;	2000US-0249265P.	CC	
PR	17-NOV-2000;	2000US-0249297P.	CC	
PR	17-NOV-2000;	2000US-0249299P.	CC	
PR	17-NOV-2000;	2000US-0249300P.	CC	
PR	01-DEC-2000;	2000US-0250160P.	CC	
PR	01-DEC-2000;	2000US-0250391P.	CC	
PR	03-DEC-2000;	2000US-0251030P.	CC	
PR	05-DEC-2000;	2000US-0251988P.	CC	
PR	05-DEC-2000;	2000US-0256719P.	CC	
PR	06-DEC-2000;	2000US-0251479P.	CC	
PR	08-DEC-2000;	2000US-0251856P.	CC	
PR	08-DEC-2000;	2000US-0251858P.	CC	
PR	08-DEC-2000;	2000US-0251859P.	CC	
PR	08-DEC-2000;	2000US-0251989P.	CC	
PR	08-DEC-2000;	2000US-0251990P.	CC	
PR	11-DEC-2000;	2000US-0254097P.	CC	
PR	05-JAN-2001;	2001US-0259678P.	CC	
XX			XX	
PA	(HUMA-) HUMAN GENOME SCI INC.		PA	
XX	Rosen CA, Barash SC, Ruben SM;		XX	
PI			PI	
XX	WPI; 2001-465557/50.		XX	
DR	N-PSDB; AAS29109.		DR	
DR			DR	
XX	Nucleic acid molecules encoding human secreted chromosomal binding proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and Parkinson's diseases and cancers.		XX	

PF 17-JAN-2001; 2001US-00764846.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX Rosen CA, Ruben SM, Barash SC;
XX WPI; 2002-690611/74.
DR N-PSDB; ABS68249.
XX Novel DNA-binding protein useful for diagnosis, prognosis, prevention and
PT treatment of immune, hyperproliferative, respiratory, cardiovascular,
PT reproductive, endocrine, gastrointestinal and neurological disorders.
XX Claim 11; SEQ ID NO 218; 225pp; English.
XX The present invention relates to a new DNA-binding protein. The invention
CC is useful in treating, preventing, diagnosing and/or prognosing
CC immunodeficiencies (e.g. B cell immunodeficiencies, severe combined
CC immunodeficiencies), autoimmune disorders (rheumatoid arthritis, multiple
CC sclerosis, diabetes mellitus), allergic reactions and conditions (e.g.

CC asthma), inflammatory conditions, graft-versus-host disease, blood-
CC related disorders (thrombosis, atherosclerosis), hyperproliferative
CC disorders (e.g. cancer), renal disorders (e.g. acute glomerulonephritis),
CC cardiovascular disorders (e.g. arrhythmic), respiratory disorders
CC (Goodpasture's syndrome), neurological disorders (e.g. Alzheimer's
CC disease, Parkinson's disease), endocrine disorders (e.g. Addison's
CC disease), reproductive system disorders (e.g. endometriosis), infectious
CC diseases (e.g. viral, bacterial or fungal infections) and
CC gastrointestinal disorders (e.g. Crohn's disease). The invention is also
CC useful to stimulate neuronal growth and treat, prevent, and/or diagnose
CC neuronal damage which occurs in certain neuronal disorders or neuro-
CC degenerative conditions. The present amino acid sequence represents a
CC human DNA-binding protein of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC <http://seqdata.uspto.gov/sequence>
XX
SQ Sequence 228 AA;
Query Match 100.0%; Score 1040; DB 5; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.7e-85;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MATETVELHKLKLAELKQECLEGLTGTGKIDLIHRLQAYLEHAEAEEDVLGDET 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 19 MATETVELHKLKLAELKQECLEGLTGTGKIDLIHRLQAYLEHAEAEEDVLGDET 78
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 61 EEEETKPIELPVKEEPPPEKTVDAAEKVKVITSEIPQTERMQKRAERFNPVLSKK 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 79 EEEETKPIELPVKEEPPPEKTVDAAEKVKVITSEIPQTERMQKRAERFNPVLSKK 138
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 121 AAPAARFGISSVPTKGLSSDNKPMVNLDKLKERAPQFGLNVSSISRKSEDEKLKKR 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 139 AAPAARFGISSVPTKGLSSDNKPMVNLDKLKERAPQFGLNVSSISRKSEDEKLKKR 198
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 181 FGIVTSAGTGTTEDTEAKKRAERFGIA 210
Db 199 FGIVTSAGTGTTEDTEAKKRAERFGIA 228
RESULT 6
ADC25371
ID ADC25371 standard; protein; 228 AA.
XX
AC ADC25371;
XX
XX
DT 18-DEC-2003 (first entry)
XX
DE Human extracellular matrix protein from gene 80.
XX
XX Extracellular matrix protein; cytostatic; antibacterial; virucide;
KW neuroprotective; gynaecological; gastrointestinal-Gen; cardiac;
KW cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;
KW respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
KW neurotropic; anti-allergic; cancer; bacterial infection; viral infection;
KW neural disorder; immune system disorder; blood disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW inflammatory disorder; proliferative disorder; human; gene therapy.
XX
OS Homo sapiens.
XX
XX US2003049650-A1.
XX
XX 13-MAR-2003.
XX
XX 07-MAR-2002; 2002US-00091483.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234937P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250191P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764846.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Ruben SM, Barash SC;
WPI; 2003-605749/57.
N-PSDB; ADC25243.
New DNA-binding proteins and gene encoding them, useful for diagnosing,
treating and/or preventing e.g. neurological, inflammatory, infectious,
cardiovascular, autoimmune, respiratory, neoplastic or digestive
diseases.
Claim 11; SEQ ID NO 218; 226pp; English.
PS

XX		The invention relates to an isolated nucleic acid molecule (cDNA)
CC		encoding a human extracellular matrix protein, representing one of 161
CC		novel genes. Also included are recombinant vectors, host cells
CC		(expressing the protein), the extracellular matrix proteins (including
CC		their fragments, epitopes and homologues), an isolated antibody that
CC		binds specifically to the protein, diagnosing a pathological condition or
CC		susceptibility to a pathological condition (comprising determining the
CC		presence or absence of a mutation in the nucleic acid and diagnosing a
CC		condition based on the presence or absence of the mutation), diagnosing a
CC		pathological condition or susceptibility to a pathological condition
CC		(comprising determining the presence or amount of expression of the
CC		protein in a biological sample and diagnosing a condition based on the
CC		presence or amount of expression of the protein), preventing, treating or
CC		ameliorating a medical condition by administering the nucleic acid or
CC		protein to a mammalian subject, identifying a binding partner to the
CC		protein, the gene corresponding to the cDNA sequence, and identifying an
CC		activity in a biological assay (comprising expressing the nucleic acid in
CC		a cell, isolating the supernatant, detecting an activity in a biological
CC		assay and identifying the protein in the supernatant having the
CC		activity). The nucleic acids and proteins display the following
CC		activities Cytostatic, antibacterial, Virucide, Neuroprotective,
CC		Gynaecological, Gastrointestinal-Gen, Cardiant, Cardiovascular-Gen,
	Query Match	100.0%; Score 1040; DB 7; Length 228;
	Best Local Similarity	100.0%; Pred. No. 1.7e-85;
	Matches 210; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	1	MATETVELHLKLAELKQECIARGLETGKIQDLIHRLOAYLEEHAEENEEDVLGDET 60
DB	19	MATETVELHLKLAELKQECIARGLETGKIQDLIHRLOAYLEEHAEENEEDVLGDET 78
QY	61	EESTKPIELPVKEEPEPKTVDVAAEKVVVKITSEIPQTERMQKRERFNVPSLESKK 120
DB	79	EESTKPIELPVKEEPEPKTVDVAAEKVVVKITSEIPQTERMQKRERFNVPSLESKK 138
QY	121	AARAARFGISSVPVTGKLSSDNKPMVNLDKLKERAQRFGLINVSSISRKSEDEKLKKRKR 180
DB	139	AARAARFGISSVPVTGKLSSDNKPMVNLDKLKERAQRFGLINVSSISRKSEDEKLKKRKR 198
QY	181	FGIWTSAGTGTTEDTEAKRKRAERFGIA 210
DB	199	FGIWTSAGTGTTEDTEAKRKRAERFGIA 228
RESULT 7		
IDP30022		
ID	ADP30022 standard; protein; 185 AA.	
XX	AC	ADP30022;
XX	AC	AC
DT	12-AUG-2004 (first entry)	
XX		
DE	Human secreted protein SEQ ID #789.	
XX		
KW	Cytostatic; Antiinflammatory; Immunosuppressive; Antibacterial; Virucide;	
KW	cancer; inflammatory; immune; human secreted protein.	
XX	Homo sapiens.	
XX		
PN	WO2004035732-A2.	
XX		
PD	29-APR-2004.	
XX		
PF	28-AUG-2003; 2003WO-US026780.	
XX		
PR	29-AUG-2002; 2002US-0406576P.	
PR	29-AUG-2002; 2002US-0406579P.	
PR	29-AUG-2002; 2002US-0406585P.	
PR	29-AUG-2002; 2002US-0406588P.	
PR	29-AUG-2002; 2002US-0406608P.	
PR	29-AUG-2002; 2002US-0406611P.	
PR	29-AUG-2002; 2002US-0406612P.	

```

XX
PS
XX
CC Claim 1; SEQ ID NO 2020; 428pp; English.
CC
CC The present invention relates to an isolated nucleic acid molecule
CC encoding a polypeptide which is believed to be cytostatic,
CC antiinflammatory, immunosuppressive, antibacterial and virucidal. The
CC composition and methods are useful for diagnosing, preventing and
CC treating diseases such as proliferative (e.g. cancer), inflammatory,
CC immune, metabolic, genetic, bacterial and viral diseases. The present
CC sequence represents a human secreted protein. The present sequence is
CC available on WIPOMEDS and is not in the specification.
XX
SQ Sequence 185 AA;

Query Match 71.7%; Score 746; DB 8; Length 185;
Best Local Similarity 85.6%; Pred. No. 4.2e-59;
Matches 160; Conservative 9; Mismatches 12; Indels 6; Gaps 4;

QY 24 GLETGKIQDLIHLRQAYLSEHAEANEEDVLGDETEETKPIELPVKEEPPPEKTVND 83
Db 5 GLETGKIQDLIHLRQAYLSEHAEANE--EDVLGDETEEEK-NPLS---NRKNPLKKTVD 59

QY 84 VAAEKVKVKITSEIPQTERMQKRAERFNVPSLESKKAARAFGISSVPTKGLSSDNKP 143
Db 60 VAAEKVKVKITSEIPQTERMQKRAEQSFVPVSLESKKAQAARFGISSVPTK-VCHLNT 118

QY 144 MVNLDKIKERAQRFGLNVSSISRSKSEDEKIKRKRERFGIVTSSAGTGTTEDTEAKKRK 203
Db 119 MVNLDKPKERAQRFGLNVSSISRSKSEDDKIKRKRERFGIVTSSAGTGTTEDTEAKKRK 178

QY 204 AERFGIA 210
Db 179 AERFGIA 185

RESULT 8
AAG00732
ID AAG00732 standard; protein; 149 AA.
XX
AC AAG00732;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 4813.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-00200610.
XX
XX 26-FEB-1999; 99US-0122487P.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
DR N-PSDB; AAC00738.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 13; SEQ ID NO 4813; 71pp + Sequence Listing; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number of
XX 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were

```

```

CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC tissues. EST sequences usually correspond mainly to the 3' untranslated
CC region (UTR) of the mRNA because they are often obtained from oligo-dT
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC sequences derived from the 5' ends of mRNAs and even in those cases where
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC are also used in diagnostic, forensic, gene therapy and chromosome
CC mapping procedures. They are used to obtain upstream regulatory sequences
CC and to design expression and secretion vectors
XX
SQ Sequence 149 AA;

Query Match 71.3%; Score 742; DB 3; Length 149;
Best Local Similarity 100.0%; Pred. No. 7.3e-59;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHLRQAYLSEHAEANEEDVLGDET 60
Db 1 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHLRQAYLSEHAEANEEDVLGDET 60

QY 61 ESEETKPIELPVKEEPPPEKTVDAAEKVKVKITSEIPQTERMQKRAERFNVPSLESK 120
Db 61 ESEETKPIELPVKEEPPPEKTVDAAEKVKVKITSEIPQTERMQKRAERFNVPSLESK 120

QY 121 AARARFGISSVPTKGLSSDNKPMVNLDK 149
Db 121 AARARFGISSVPTKGLSSDNKPMVNLDK 149

RESULT 9
ABB63720
ID ABB63720 standard; protein; 308 AA.
XX
AC ABB63720;
XX
DT 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 17952.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL07823.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 17952; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention

```



```

XX 02-OCT-2002; 2002US-0416186P.
XX (NUVE-) NUVELO INC.
XX
XX Tang YT, Asundi V, Ren F, Zhang J, Wehrman T, Wang Z, Ma Y;
XX Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;
XX WPI; 2004-668857/65.
XX N-PSDB; ADS09800.
XX
XX New polynucleotide, useful in preparing a composition for diagnosing or
XX treating inflammatory, neurodegenerative or stem cell disorders, e.g.,
XX aplastic anemia or cancer for promoting wound healing.
XX
XX Claim 20; SEQ ID NO 721; 718pp; English.
XX
XX The invention relates to a novel isolated polynucleotide and the encoded
XX polypeptide. The molecules of the invention demonstrate antiinflammatory,
XX neuroprotective, antianaemic, cytostatic and vulnerary activities and may
XX be useful in preparing a composition for diagnosing or treating
XX inflammatory, haematopoietic, immune, neurodegenerative or stem cell
XX disorders, such as aplastic anaemia or cancer, as well as for promoting
XX wound healing. The molecules may also be utilised during gene therapy
XX procedures. The current sequence is that of a human therapeutic protein
XX of the invention. The current sequence is not shown explicitly within the
XX specification but can be accessed from the WIPO web-site.
XX
XX Sequence 747 AA;
Query Match 14.9%; Score 155; DB 8; Length 747;
Best Local Similarity 26.2%; Pred. No. 4.5e-05;
Matches 61; Conservative 37; Mismatches 77; Indels 58; Gaps 9;
QY 6 VELHKLKLAELKQFCILARGLETGKIKODLTHRLQAYLE-EHAE----- 50
DB 1 MEVKRLKVTLSLQRLQRLGSLKVDLAQLQALDAEMLEDEAGGGGAGPGACKAE 60
QY 51 -----NEEDVLGDETEETKPIELPVKEEPEPEKTVDAAEKVKVITSEI 97
DB 61 PRVAAGGGPGGDEEB---DEEEEDDEBALLEDEDEPP-----PAQALGQAAQPPPEP 113
QY 98 POTERMOKRAERNVPSLESKKAARAARFGISSVPTKGLSSDNKPMVNLDKLKERAA--Q 155
DB 114 PEAAEAAAE---PDAEKPAAETAGSGVNGEEOGLKREE-----DEPEERSGDE 164
QY 156 RFLGNV--SSISRKSEDEKLKKRKRFGITVTSAGTGTTEDEAKKRAER 206
DB 165 TPGSEVPGDKAAEEQGDQDSEKSP-----AGSDGERRGVKQRDEK 207
XX
XX RESULT 12
XX ABB61567
XX ID ABB61567 standard; protein; 795 AA.
XX AC ABB61567;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 11493.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL05670.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 11493; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 795 AA;
Query Match 14.9%; Score 155; DB 4; Length 795;
Best Local Similarity 21.0%; Pred. No. 4.9e-05;
Matches 55; Conservative 49; Mismatches 86; Indels 72; Gaps 8;
QY 4 ETVELHKLKLAELKQFCILARGLETGKIKODLTHRLQAYLE-----EHA----- 46
DB 2 DVAKLEKMKVLDRLNELQSLGSLDTGKGAVLVERLAYVEGGAGDGENAPVTPSRQRRT 61
QY 47 -----EEEAENEEDVLGDETEETKPIELPVKEEPEPEKTV 82
DB 62 RSMRSFSPVQAAPVAAPVILDTLEEEQEDKTVQPEPESEQPAAEPEPESEPEAE 121
QY 83 DVAAEKVKVITSEI-----POTERMOKRAERNVPSLESKKAARAARFGISSVPTK-- 135
DB 122 PAAAVTDTTVAQVNVNEESQPEPEFDEKSETDDKQETIEEAVPAVVPQNEVADEPMEED 181
QY 136 --GLSSDNKPMVNLDKLKERAFGLNVSSIS--RKSEDEKLKKRKRFGITVTSAGTG 191
DB 182 HDAAPEEQEPTQTEEPVEEKPAE-----STVAEHQSGNGDSQKMDVDEE-----DSAAPK 230
QY 192 TTEDTE-----AKKRA 204
DB 231 TAETEPAAKPEDQPPERRKRS 252
XX
XX RESULT 13
XX AAG20995
XX ID AAG20995 standard; protein; 185 AA.
XX AC AAG20995;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 23391.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX

```


PF 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-01231180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0128234P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139751P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 18-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.

PR	28-SEP-1999;	99US-0156458P.	XX	25-FEB-2000;	2000EP-00301439.	XX
PR	29-SEP-1999;	99US-0156596P.	PF			
PR	04-OCT-1999;	99US-0157117P.	XX			
PR	05-OCT-1999;	99US-0157753P.	PR	25-FEB-1999;	99US-0121825P.	
PR	06-OCT-1999;	99US-0157865P.	PR	09-MAR-1999;	99US-0123180P.	
PR	07-OCT-1999;	99US-0158029P.	PR	09-MAR-1999;	99US-0123548P.	
PR	08-OCT-1999;	99US-0158232P.	PR	23-MAR-1999;	99US-0125788P.	
PR	12-OCT-1999;	99US-0158369P.	PR	25-MAR-1999;	99US-0126264P.	
PR	13-OCT-1999;	99US-0159293P.	PR	29-MAR-1999;	99US-0126785P.	
PR	13-OCT-1999;	99US-0159294P.	PR	01-APR-1999;	99US-0127462P.	
PR	13-OCT-1999;	99US-0159295P.	PR	06-APR-1999;	99US-0128234P.	
PR	14-OCT-1999;	99US-0159329P.	PR	08-APR-1999;	99US-0128714P.	
PR	14-OCT-1999;	99US-0159330P.	PR	16-APR-1999;	99US-0129845P.	
PR	14-OCT-1999;	99US-0159331P.	PR	19-APR-1999;	99US-0130077P.	
PR	14-OCT-1999;	99US-0159637P.	PR	21-APR-1999;	99US-0130449P.	
PR	14-OCT-1999;	99US-0159638P.	PR	23-APR-1999;	99US-0130510P.	
PR	21-OCT-1999;	99US-0159584P.	PR	23-APR-1999;	99US-0130891P.	
PR	21-OCT-1999;	99US-0160741P.	PR	28-APR-1999;	99US-0131449P.	
PR	21-OCT-1999;	99US-0160767P.	PR	30-APR-1999;	99US-0132048P.	
PR	21-OCT-1999;	99US-0160768P.	PR	30-APR-1999;	99US-0132407P.	
PR	21-OCT-1999;	99US-0160770P.	PR	04-MAY-1999;	99US-0132484P.	
PR	21-OCT-1999;	99US-0160814P.	PR	05-MAY-1999;	99US-0132485P.	
PR	21-OCT-1999;	99US-0160815P.	PR	06-MAY-1999;	99US-0132486P.	
PR	22-OCT-1999;	99US-0160980P.	PR	06-MAY-1999;	99US-0132487P.	
PR	22-OCT-1999;	99US-0160981P.	PR	07-MAY-1999;	99US-0132487P.	
PR	22-OCT-1999;	99US-0160989P.	PR	11-MAY-1999;	99US-0132863P.	
PR	25-OCT-1999;	99US-0161404P.	PR	11-MAY-1999;	99US-0134256P.	
PR	25-OCT-1999;	99US-0161405P.	PR	14-MAY-1999;	99US-0134218P.	
PR	25-OCT-1999;	99US-0161406P.	PR	14-MAY-1999;	99US-0134219P.	
PR	26-OCT-1999;	99US-0161359P.	PR	14-MAY-1999;	99US-0134221P.	
PR	26-OCT-1999;	99US-0161360P.	PR	14-MAY-1999;	99US-0134370P.	
PR	26-OCT-1999;	99US-0161361P.	PR	18-MAY-1999;	99US-0134768P.	
PR	26-OCT-1999;	99US-0161362P.	PR	19-MAY-1999;	99US-0134941P.	
PR	28-OCT-1999;	99US-0161920P.	PR	20-MAY-1999;	99US-0135124P.	
PR	28-OCT-1999;	99US-0161952P.	PR	21-MAY-1999;	99US-0135353P.	
PR	28-OCT-1999;	99US-0161993P.	PR	24-MAY-1999;	99US-0135629P.	
PR	29-OCT-1999;	99US-0162142P.	PR	25-MAY-1999;	99US-0136021P.	
Query Match 13.4%; Score 139; DB 3; Length 185;						
Best Local Similarity 31.2%; Pred. No. 0.00021;						
Matches 55; Conservative 22; Mismatches 67; Indels 32; Gaps 7;						
Qy	57	GDETEEBETPIELPVKEEPPPEKTVD-----VAAEKVKVITGEIPOT----- 100				
Db	17	GLSTGENPKIVDLNLTETLDRDILDGEVKGFSDSGEKEETDSNGIGTAGVDGSDIS 76				
Qy	101	-----ERMQKRAERFNPVSL--ESKKAARAARFGISSVPTKGLSSDNKPMVNLDKLKER 153				
Db	77	PVDDIQKKIRRAERFGVSVKLTTEEKRNRAERFGTVAAAVVNGSEGTAKAEL-KRKAR 135				
Qy	154	AQRFGI-NVSSI3RKSDEDEKLKKRERFGIVTSSAGTGTTEAKRK-RAERF 207				
Db	136	ADRFVPSATSTTDKTEEEAKKARLARFGKETK-----VDSAEENKKRARALRF 185				
RESULT 14						
AAG22070						
ID	AAG22070 standard; protein; 214 AA.					
XX	AC					
XX	AAG22070;					
DT	17-OCT-2000 (first entry)					
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 24859.					
DE	XX					
KW	Protein identification; signal transduction pathway; metabolic pathway;					
KW	hybridisation assay; genetic mapping; gene expression control; promoter;					
KW	termination sequence.					
XX	XX					
OS	Arabidopsis thaliana.					
XX	XX					
PN	EP1033405-A2.					
XX	XX					
PD	06-SEP-2000.					

PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145991P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149436P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.

PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161952P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 13.4%; Score 139; DB 3; Length 214;
Best Local Similarity 31.2%; Pred. No. 0.00026;
Matches 55; Conservative 22; Mismatches 67; Indels 32; Gaps 7;

Qy 57 GDETEEBETPIELPVKEEPEPKTVD-----VAAEKVKVITSEIPOT----- 100
Db 17 GLSTGENPKKIVDLNTTTELDRTDDILDGEVKGFSDSGEKKEBTDSDNGIGSTAGVDSGDIS 76
Qy 101 -----ERMQKRAERNVPVSL--ESKKAARAARFGISSVPTKGLSSDNKPMVLDLKER 153
Db 77 PVDDIQKKIRRAERFGVSVKLTTEERKNSRAERFGTVAAAVVNGSEGTAKBEL-KRKAR 135
Qy 154 AQRFGGL-NVSSISRKSDDEKLKKRKERFGIVTSSAGTGTTEDEAKRK-RARERF 207
Db 136 ADRFGVPSATSTTDKTEEEAKKARLARFGKETK-----VDSAEENKRRARALRF 185

RESULT 15

ADRO8732
ID ADRO8732 standard; protein; 750 AA.

XX AC ADRO8732;

XX DT 04-NOV-2004 (first entry)

DE Human protein useful for treating neurological disease Seq 2238.

XX human; oligo-capping method; diagnostic marker; gene therapy;
KW osteoporosis; neurological disease; Alzheimer's disease;
KW Parkinson's disease; dementia; short memory; cancer;
KW sense or motor function; emotional reaction; fear response; panic;
KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;
KW tranquiliser.

XX OS Homo sapiens.

```
XX EP1447413-A2.
PN
XX
XX
PD 18-AUG-2004.
XX
XX 12-FEB-2004; 2004EP-00003145.
XX
XX 14-FEB-2003; 2003JP-00102207.
PR
XX 09-MAY-2003; 2003JP-00131452.
PR
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
PA
XX
XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI Wakamatsu A, Ishii S, Nagai K, Irie R;
PI
XX
DR WPI; 2004-583265/57.
DR N-PSDB; ADR06776.
XX
XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
PT
XX Claim 1; SEQ ID NO 238; 2686pp; English.
PS
XX This invention relates to novel, isolated full length human cDNA
CC molecules and the encoded proteins thereof. Specifically, it refers to
CC cDNA clones obtained by an oligo-capping method, where none of these
CC clones are identical to any known human mRNAs. The present invention
CC describes an immunoassay to identify agonists and antagonists, as well as
CC antibodies, antisense molecules and siRNAs that can all be used to bind
CC to and modulate expression of the cDNA molecules. As such, these
CC molecules are useful for diagnostic markers or therapeutic targets for
CC the various diseases or morbid states. In particular, they are useful in
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC disease, Parkinson's disease, dementia, short memory and various cancers,
CC as well as for maintaining equilibrium of sense or motor function, and
CC for treating emotional reaction, fear response and panic. Accordingly,
CC they exhibit osteopathic, neuroprotective, neurotropic, antiparkinsonian,
CC cytotatic and tranquiliser activities. This polypeptide is a protein
CC encoded by a full length human cDNA sequence of the invention. NOTE: This
CC sequence is not given in the sequence listing of the specification but
CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC office.
XX
XX Sequence 750 AA;
SQ
Query Match 13.3%; Score 138; DB 8; Length 750;
Best Local Similarity 25.4%; Pred. No. 0.0015;
Matches 60; Conservative 27; Mismatches 77; Indels 72; Gaps 9;
QY 1 MATETVELHKLKLAELKQECIARGLETGKIKQDLIHLQAYLEH-----AEEAN--- 51
Db 1 MSSSPVNVKKLKVSELKELKKRLSDGKLKAEIWMERLQALDDDEAGRPAPWPGNGDD 60
QY 52 -----EDVLGDETE-----EETKPIELPVKEEPEPEKTVDVAAE-----KK 89
Db 61 QGFQEGEDELGDDEEGAGDENGHEGQQPQ---PATQQQPPQQQGAKEAGKSSGPTS 117
QY 90 VVKITSIPQTERNQKKAERFNPVSVLESKKAARAARFGISSVPTKGLSSDNKPMVNLK 149
Db 118 LFAVTVAPPGARQQQAG-----GKKKAGGGGGGRPGAPAG---DGK----- 159
QY 150 LKERAQRFGLNVSSISRKSEDDEKLKKERFGIVTSSAGTGTTEAKKRAE 205
Db 160 ---TEQGGDKKRGVKRPREDH-----GRGYFEYIENKYSRAK 195
```

Search completed: October 5, 2005, 05:57:27
Job time : 108 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 05:59:14 ; Search time 1324 Seconds
(without alignments)
65.842 Million cell updates/sec

Title: US-09-788-476A-2

Perfect score: 1040

Sequence: 1 MATETVELHKLKLAELKQEC.....GTTEDTEAKRKRAERFGIA 210

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1846076 seqs, 415116000 residues

Total number of hits satisfying chosen parameters: 1846076

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1040	100.0	210	9	US-09-788-476A-2
2	1040	100.0	228	9	US-09-764-846-218
3	1040	100.0	228	14	US-10-081-483-218
4	222	21.3	308	20	US-11-097-143-17952
5	160	15.4	169	16	US-10-425-115-335173
6	155	14.9	734	16	US-10-483-506-6
7	155	14.9	795	20	US-11-097-143-11493
8	146.5	14.1	165	15	US-10-424-599-156346
9	146	14.0	115	15	US-10-425-114-48886
10	143	13.8	175	16	US-10-437-963-176147
11	140	13.5	153	16	US-10-767-701-32849

12	138.5	13.3	189	15	US-10-424-599-156348	Sequence 156348,
13	138.5	13.3	212	15	US-10-425-114-44717	Sequence 44717, A
14	137.5	13.2	540	14	US-10-106-698-5766	Sequence 5766, Ap
15	137.5	13.2	1150	15	US-10-433-544-4	Sequence 4, Appli
16	136.5	13.1	1150	14	US-10-316-532-4	Sequence 4, Appli
17	136.5	13.1	1172	17	US-10-828-815-35	Sequence 35, Appli
18	133	12.8	1146	14	US-10-316-532-2	Sequence 2, Appli
19	133	12.8	1153	17	US-10-828-815-33	Sequence 33, Appli
20	124	11.9	775	10	US-09-934-455-450	Sequence 450, App
21	122.5	11.8	805	15	US-10-369-493-152	Sequence 152, App
22	122	11.7	2274	20	US-11-097-143-2763	Sequence 2763, Ap
23	120.5	11.6	450	15	US-10-094-749-2630	Sequence 2630, Ap
24	120.5	11.6	824	20	US-11-009-554-23	Sequence 23, Appli
25	120.5	11.6	2020	15	US-10-369-493-5128	Sequence 5128, Ap
26	120.5	11.6	2020	15	US-10-369-493-5129	Sequence 5129, Ap
27	119.5	11.5	686	16	US-10-425-115-213836	Sequence 213836,
28	119.5	11.5	1332	9	US-09-982-091A-4	Sequence 4, Appli
29	119	11.4	268	15	US-10-412-699B-730	Sequence 730, App
30	119	11.4	811	15	US-10-264-049-3051	Sequence 3051, Ap
31	118.5	11.4	845	15	US-10-205-331-61	Sequence 61, Appli
32	118	11.3	303	16	US-10-425-115-213843	Sequence 213843,
33	116.5	11.2	36946	18	US-10-840-512-155	Sequence 155, App
34	116	11.2	713	16	US-10-437-963-177006	Sequence 177006,
35	116	11.2	806	16	US-10-684-422-184	Sequence 184, App
36	116	11.2	806	20	US-11-009-554-22	Sequence 22, Appli
37	115.5	11.1	856	18	US-10-764-425-154	Sequence 154, App
38	115.5	11.1	1150	20	US-11-097-143-4179	Sequence 4179, Ap
39	115	11.1	428	14	US-10-254-995-7	Sequence 7, Appli
40	115	11.1	428	20	US-11-062-080-7	Sequence 7, Appli
41	115	11.1	439	10	US-09-056-019-37	Sequence 37, Appli
42	115	11.1	439	16	US-10-751-702-37	Sequence 37, Appli
43	115	11.1	439	20	US-11-032-644-37	Sequence 37, Appli
44	115	11.1	1359	17	US-10-732-923-8707	Sequence 8707, Ap
45	115	11.1	1359	17	US-10-732-923-8708	Sequence 8708, Ap

ALIGNMENTS

RESULT 1

US-09-788-476A-2
; Sequence 2, Application US/09788476A
; Patent No. US20020107190A1
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Ching M. et al.
; TITLE OF INVENTION: NOVEL GENES AND EXPRESSION PRODUCTS THEREFROM
; FILE REFERENCE: 3669-0103P
; CURRENT APPLICATION NUMBER: US/09/788,476A
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 210
; TYPE: PRT
; ORGANISM: human
US-09-788-476A-2

Query Match 100.0%; Score 1040; DB 9; Length 210;
Best Local Similarity 100.0%; Pred. No. 2.9e-76;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MATETVELHKLKLAELKQEC	LARGLETGKIGKQDLIHLRQAYLEHAEANEEDVLGDET	60
Db	1	MATETVELHKLKLAELKQEC	LARGLETGKIGKQDLIHLRQAYLEHAEANEEDVLGDET	60
Qy	61	EEETKPIELPVKEEPEPKTV	VAABKVKVKTITSETPQTERMQRAERFNVPSLESKK	120
Db	61	EEETKPIELPVKEEPEPKTV	VAABKVKVKTITSETPQTERMQRAERFNVPSLESKK	120
Qy	121	AAARAARFGISVPPTKGLSS	DNKPMVNDKUKERAQRLNVSISRKSEDEKLUKKRER	180
Db	121	AAARAARFGISVPPTKGLSS	DNKPMVNDKUKERAQRLNVSISRKSEDEKLUKKRER	180

```
Qy 181 FGIWSSAGTGTEDTEAKKRAERFGIA 210
|||||
Db 181 FGIWSSAGTGTEDTEAKKRAERFGIA 210
|||||

RESULT 2
US-09-764-846-218
; Sequence 218, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ12
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 218
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-846-218

Query Match 100.0%; Score 1040; DB 9; Length 228;
Best Local Similarity 100.0%; Pred. No. 3.2e-76;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAFEEANEEDVLGDET 60
|||||
Db 19 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAFEEANEEDVLGDET 78
|||||

Qy 61 EBEETKPIELPVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERNVPSLESKK 120
|||||
Db 79 EBEETKPIELPVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERNVPSLESKK 138
|||||

Qy 121 AAAAAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
|||||
Db 139 AAAAAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 198
|||||

Qy 181 FGIWSSAGTGTEDTEAKKRAERFGIA 210
|||||
Db 199 FGIWSSAGTGTEDTEAKKRAERFGIA 228
|||||

RESULT 3
US-10-091-483-218
; Sequence 218, Application US/10091483
; Publication No. US20030049650A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ12C1
; CURRENT APPLICATION NUMBER: US/10/091,483
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 348
; Prior application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 218
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-091-483-218

Query Match 100.0%; Score 1040; DB 14; Length 228;
Best Local Similarity 100.0%; Pred. No. 3.2e-76;
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAFEEANEEDVLGDET 60
|||||
Db 19 MATETVELHKLKLAELKQECCLARGLETGKIQDLIHRLOAYLEHAFEEANEEDVLGDET 78
|||||

Qy 61 EBEETKPIELPVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERNVPSLESKK 120
|||||
```

```
Db 79 EBEETKPIELPVKEEPEPEKTVDAAEKKVVKITSEIPQTERMQKRAERNVPSLESKK 138
|||||
Qy 121 AAAAAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 180
|||||
Db 139 AAAAAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSEDEKLKKRKR 198
|||||
Qy 181 FGIWSSAGTGTEDTEAKKRAERFGIA 210
|||||
Db 199 FGIWSSAGTGTEDTEAKKRAERFGIA 228
|||||

RESULT 4
US-11-097-143-17952
; Sequence 17952, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17952
; LENGTH: 308
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-17952

Query Match 21.3%; Score 222; DB 20; Length 308;
Best Local Similarity 27.3%; Pred. No. 7.7e-10;
Matches 83; Conservative 40; Mismatches 77; Indels 104; Gaps 10;

Qy 7 ELHKLKLAELKQECCLARGLETGKIQDLIHRLOA-----YLEEHAEEANEED----- 54
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 5 DVTMKVADLKRELKRLGLAVNGKTELQDLRLOTALLEGDLSEDSAIADAIDDDVSFT 64
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 55 -----VLGDETEEE-----ETKPIEL----- 70
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 65 DEDEHKLGDENDELLKSPVSTPTTVAIPDLIAEEKTSAPDAAAPTKKIVLKRNSQ 124
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 71 -----PVKEEPEP-----EKTVDV-----AAEKKVVKITS 95
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 125 STGTVASTGTTTSPKNEAPAAAASDSTGETPTKKHKPIVVGPKTEGKPSGDKKLNQITA 184
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 96 EIPQTERMQKRAERNV-PVLSLESKKAARAARFGISSVPTKGLSSDNKPMVNLDK--LKE 152
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 185 Q-----ERLEIRAKKFGITTPAVANTATAVAINKSSASITANKGNKETEEQKEALKK 240
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 153 RQRFGLNVSSISRKSEDEKLKKRERFGIIVTSAGTGTEDTEA-----KKRKAER 206
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 241 RAERFGVVPDKAPTSKADRLQKRERFGAGAVSAATTTPTTTTESKDAWSEKARARLER 300
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

```
Qy 207 FGIA 210
Db 301 FKTA 304

RESULT 5
US-10-425-115-335173
; Sequence 335173, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 363326
; SEQ ID NO 335173
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_68790C.1.pap
US-10-425-115-335173

Query Match 15.4%; Score 160; DB 16; Length 169;
Best Local Similarity 34.0%; Pred. No. 3.8e-05;
Matches 53; Conservative 18; Mismatches 63; Indels 22; Gaps 6;

Qy 59 ETEBEETKPIELPVKEBEPPEKTVDAEKKVKITSEIPQT--ERMQKRAERFNPV-- 114
Db 7 KTADTAAPPPVAVTGEAAPGASSTPAQNPENAAATAVAGATDLEKKVRAERFGMPVLM 66

Qy 115 SLESKKAARAARFGISSVPTKGLSSDNKPMVNLKLERQAFGLNVSSISRKSEDEKL 174
Db 67 SESEKRSRAERFGTSSSVK-----EEKKKRAERFGLASPS---SSDEAKK 113

Qy 175 KKRERFGIVTSSAGTGTTEDEAKKRAERFGIA 210
Db 114 KARLERFG---QSANVDKAE--EEKKARAARFAEA 144

RESULT 6
US-10-483-506-6
; Sequence 6, Application US/10483506
; Publication No. US20040249127A1
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Becha, Shanya D.
; APPLICANT: Chim, Anna M.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Ding, Li
; APPLICANT: Forsythe, Ian J.
; APPLICANT: Gietzen, Kimberly J.
; APPLICANT: Griffin, Jennifer A.
; APPLICANT: Hafalia, April J. A.
; APPLICANT: Khan, Farrah A.
; APPLICANT: Lal, Preeti G.
; APPLICANT: Lee, Ernestine A.
; APPLICANT: Li, Joana X.
; APPLICANT: Lu, Dyung Aina M.
; APPLICANT: Ramkumar, Jayalaxmi
; APPLICANT: Richardson, Thomas W.
; APPLICANT: Swarnakar, Anita
; APPLICANT: Tang, Y. Tom
; APPLICANT: Walia, Narinder K.
; APPLICANT: Warren, Bridget A.
; APPLICANT: Yang, Junming
; APPLICANT: Yao, Monique G.
```

```
; APPLICANT: Yue, Henry
; APPLICANT: Zebardjian, Yeganeh
; TITLE OF INVENTION: Intracellular Signaling Molecules
; FILE REFERENCE: PF-1066 USN
; CURRENT APPLICATION NUMBER: US/10/483,506
; CURRENT FILING DATE: 2004-01-12
; PRIOR APPLICATION NUMBER: PCT/US02/22379
; PRIOR FILING DATE: 2002-07-11
; PRIOR APPLICATION NUMBER: US 60/305,113
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US 60/305,367
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/306,966
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 60/308,175
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/308,327
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/309,902
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/310,752
; PRIOR FILING DATE: 2001-08-07
; PRIOR APPLICATION NUMBER: US 60/311,636
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 1910008CD1
US-10-483-506-6

Query Match 14.9%; Score 155; DB 16; Length 734;
Best Local Similarity 26.2%; Pred. No. 0.00099;
Matches 61; Conservative 37; Mismatches 77; Indels 58; Gaps 9;

Qy 6 VELHLKLAELKQECIARGLETGKIQDLIHLRQAVLE-EHABEEA----- 50
Db 1 MEVKRLKVTLSRSELQRRGLDSRGLKVDLAQRLEALDAEMLEDEAGGGAGPGACKAE 60

Qy 51 -----NEEDVLGDETEEBETKPIELPVKEEPEKTVDAEKKVKITSEI 97
Db 61 PRPVAASGGPGGDEBE---DEEBEEDDEALLEDEDEPP-----PAQALGQAAQPPPEP 113

Qy 98 POTERMQKRAERFNPVSVLSKKAARAARFGISSVPTKGLSSDNKPMVNLKLERQAFGLNVSSISRKSEDEKL 174
Db 114 PEAAAMEAAAE---PDASEKPAEATAGSGGVNGBEQGLGKREE---DEPEERSGDE 164

Qy 156 RFGLNV--SSISRSKSEDEKLKKRERFGIVTSSAGTGTTEDEAKKRAER 206
Db 165 TPGSEVPDGKAAEEQGDQDSEKSP-----AGSDGERRGVKQRDEK 207

RESULT 7
US-11-097-143-11493
; Sequence 11493, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
```

```
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11493
; LENGTH: 795
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-11493

Query Match          14.9%; Score 155; DB 20; Length 795;
Best Local Similarity 21.0%; Pred. No. 0.00065;
Matches 55; Conservative 49; Mismatches 86; Indels 72; Gaps 8;

QY 4 ETVELHKLAEKQECLEGLTKGIKODLIHRLQAYLE-----EHA----- 46
DB 2 DVAKLEKMKVVDLRLNELQSLGLTKGVKVLVERLRAVVEGGAGDGENAPVTPSRQRRT 61
QY 47 -----EEANEEDVLGDTEETETKPIELPVKEEPEPEKT 82
DB 62 RMSRSPSPVQAAAPVAAPVLDLLEEEQEDTKVPQPESEQPAAPPEPEQSEPEEAE 121
QY 83 DVAEAKKVKITSEI-----PQTERMQKRAERFNVPSLESKKAARAARFGISVTK-- 135
DB 122 PAAAVTEDTTVNOAVNESQPEPEEFDEKSETDDKQETIEEAVPAVVPQNEVADEPMEED 181
QY 136 --GLSSDNKPMVLDLKLKERAQFGLNVSSIS--RKSEDDKLLKRRKRGIVTSSAGTG 191
DB 182 HDAAPQEQTPTQTEEPVEEKPAE-----STVAEHQNSGDSQKMDVDEE-----DSAAPK 230
QY 192 TTEDTE-----AKKRRKA 204
DB 231 TABETPAAKPEQPQPPERRKRS 252

RESULT 8
US-10-424-599-156346
; Sequence 156346, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156346
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112200C.1.pap
US-10-424-599-156346

Query Match          14.1%; Score 146.5; DB 15; Length 165;
Best Local Similarity 31.1%; Pred. No. 0.00046;
Matches 42; Conservative 26; Mismatches 46; Indels 21; Gaps 5;

QY 75 EEPPEKTVDAEAKKVKVKITSEIPQTERMQKRAERFNVPSLESKKAARAARFGISSVPT 134
```

```
DB 13 QENPNKTL-D-PTPEPDPDIPATEDATATK-----GSDAKDPADADADAITSPPP 58
QY 135 KGLSSDNKPMVLDLKLKERAQFGLNVSSISRKSEDDKLLKRRKRGIVTSSAGTGTTT 194
DB 59 D--SGNDAPFLSDIQKMRRAERFGISV-QLSEK-----EKRNSRAERFGTVSASQSEPSK 111
QY 195 DTEAKKRRKRAERFGI 209
DB 112 SEDLKRKARAERFGM 126

RESULT 9
US-10-425-114-48886
; Sequence 48886, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48886
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3912-045-B5_FLI.pap
US-10-425-114-48886

Query Match          14.0%; Score 146; DB 15; Length 115;
Best Local Similarity 40.7%; Pred. No. 0.00033;
Matches 44; Conservative 12; Mismatches 32; Indels 20; Gaps 5;

QY 105 KRAERFNVFV--SLESKKAARAARFGISSVPTKGLSSDNKPMVLDLKLKERAQFGLNV 162
DB 1 RRAERFGMPVLMSEEEKSRRAERFGTGSSVK-----EEKKKRAERFGLASP 50
QY 163 SISRKSEDDKLLKRRKRGIVTSSAGTGTTTTEAKKRRARFGIA 210
DB 51 S---SSDEEAKKARLERFG---QSANVDKAE--EEKKARAARFAEA 90

RESULT 10
US-10-437-963-176147
; Sequence 176147, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 176147
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Oryza sativa
```



```
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_73924C.1.pap
US-10-437-963-176147

Query Match      13.8%; Score 143; DB 16; Length 175;
Best Local Similarity 39.4%; Pred. No. 0.00095;
Matches 43; Conservative 14; Mismatches 32; Indels 20; Gaps 5;

Qy 101 ERMQKRAERFN--VPVSLESKKAARARFGISSVPTKGLSSDNKPMVNLDKLKERAQRF 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 EKMRRARERFCTAVVMSEERSSRAERFGTG-----SSNEK--ABEQKKRSRAERFG 108

Qy 159 LNVSSISRKSDDEKLKKRKERFGIVTSSAGTGTEDTEAKRKRAERF 207
   || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 LASSS----AEDAKKXKARLERFG-----QSTNVDKGEERKARALRF 148

RESULT 11
US-10-767-701-32849
; Sequence 32849, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 32849
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(153)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C19564_1.pap
US-10-767-701-32849

Query Match      13.5%; Score 140; DB 16; Length 153;
Best Local Similarity 38.1%; Pred. No. 0.0014;
Matches 40; Conservative 12; Mismatches 31; Indels 22; Gaps 4;

Qy 101 ERMQKRAERFNVPV--SLESKKAARARFGISSVPTKGLSSDNKPMVNLDKLKERAQRF 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 EKMRRARERFCTPVVMEERKSRARERFGTSSSVK-----EEEKKSRABKIG 102

Qy 159 LNVSSISRKSDDEKLKKRKERFGIVTSSAGTGTEDTEAKRK 203
   || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 103 LASSS----SSDEAKKXKARLERF-----QGTNGDKAKKKKK 137

RESULT 12
US-10-424-599-156348
; Sequence 156348, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156348
; LENGTH: 189
```

```
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112202C.1.pap
US-10-424-599-156348

Query Match      13.3%; Score 138.5; DB 15; Length 189;
Best Local Similarity 30.7%; Pred. No. 0.0024;
Matches 46; Conservative 20; Mismatches 53; Indels 31; Gaps 5;

Qy 63 EETKPIELPVKEEPEPEKTVDVAAEKKVVKITSEIPQTERMQKRAERFNVPSLESKAA 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 DPTPEDPDPATEDDAADTKDSAK-----NSPPPDGSGN-----DAPLSDIQKKMR 66

Qy 123 RAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSDDEKLKK--RKER 180
   || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 RAERFGIS-----VOLSEKERNRAERFGTGSASQGSSEPSKSELKRAER 115

Qy 181 FGIVTSSAGTGTEDTEAKRKRAERFGIA 210
   || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 FGMPSPSPT-----TTSDEEAKKKARLARFAPA 141

RESULT 13
US-10-425-114-44717
; Sequence 44717, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 44717
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700751810_FLI.pap
US-10-425-114-44717

Query Match      13.3%; Score 138.5; DB 15; Length 212;
Best Local Similarity 30.7%; Pred. No. 0.0028;
Matches 46; Conservative 20; Mismatches 53; Indels 31; Gaps 5;

Qy 63 EETKPIELPVKEEPEPEKTVDVAAEKKVVKITSEIPQTERMQKRAERFNVPSLESKAA 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 44 DPTPEDPDPATEDDAADTKDSAK-----NSPPPDGSGN-----DAPLSDIQKKMR 89

Qy 123 RAARFGISSVPTKGLSSDNKPMVNLDKLKERAQRFGLNVSSISRKSDDEKLKK--RKER 180
   || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 RAERFGIS-----VOLSEKERNRAERFGTGSASQGSSEPSKSELKRAER 138

Qy 181 FGIVTSSAGTGTEDTEAKRKRAERFGIA 210
   || : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 FGMPSPSPT-----TTSDEEAKKKARLARFAPA 164

RESULT 14
US-10-106-698-5766
; Sequence 5766, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 02:29:12 ; Search time 26 Seconds
(without alignments)
777.135 Million cell updates/sec

Title: US-09-788-476A-2
Perfect score: 1040
Sequence: 1 MATETVELHLKLAELKQEC.....GTTEDTEAKKRAERFGIA 210

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1040	100.0	210	2 JC7830	cytokine-inducible
2	222	21.3	245	2 T41291	conserved hypothet
3	148	14.2	344	2 T48827	hypothetical prote
4	139	13.4	214	2 T48298	hypothetical prote
5	139	13.4	1222	2 T22490	hypothetical prote
6	135.5	13.0	3488	2 T34418	hypothetical prote
7	133.5	12.8	633	2 T05005	hypothetical prote
8	125.5	12.1	218	2 S50566	hypothetical prote
9	125	12.0	1110	2 T31116	NP-180 - sea lampr
10	123.5	11.9	849	2 S00030	neurofilament trip
11	123	11.8	363	2 F87754	protein C43E11.1 l
12	122.5	11.8	399	2 E71169	hypothetical prote
13	122.5	11.8	805	2 E70474	translation initia
14	120.5	11.6	2020	2 T21174	hypothetical prote
15	119.5	11.5	1087	2 T30330	gelsolin-related p
16	119.5	11.5	1224	2 F96795	hypothetical prote
17	119	11.4	415	2 T32490	hypothetical prote
18	118.5	11.4	517	1 A39038	l-caldesmon, nonmu
19	118.5	11.4	845	2 A45669	neurofilament trip
20	117	11.2	276	2 A25345	troponin T, cardia
21	116.5	11.2	263	2 C31957	troponin T, skelet
22	116.5	11.2	644	2 S5395	neurofilament prot
23	116	11.2	318	2 C94651	hypothetical prote
24	116	11.2	806	2 S22765	heterogeneous ribo
25	115.5	11.1	856	2 T13159	E1B-55kDa-associat
26	115.5	11.1	1877	2 T21861	hypothetical prote
27	115	11.1	1359	2 T34036	hypothetical prote
28	115	11.1	7962	2 T38346	elastic titin - hu
29	114.5	11.0	763	2 T08929	hypothetical prote

30	114	11.0	325	2 T18283	hypothetical prote
31	114	11.0	419	2 G75062	probable flagella-
32	113.5	10.9	573	2 S66710	probable membrane
33	113.5	10.9	683	2 E86358	F12K8.13 protein -
34	113.5	10.9	720	2 T26819	hypothetical prote
35	112.5	10.8	1408	2 T45039	hypothetical prote
36	111.5	10.7	718	2 AB1258	hypothetical prote
37	110.5	10.6	375	2 S26059	probable transform
38	110.5	10.6	407	2 S23325	M2 protein precurs
39	110.5	10.6	2017	1 A36014	myosin heavy chain
40	110.5	10.6	2057	2 S61477	myosin II heavy ch
41	110	10.6	771	1 A33430	h-caldesmon - chic
42	110	10.6	1603	2 S17983	gene posterior sex
43	110	10.6	1701	2 T09127	probable erythrocy
44	109.5	10.5	1156	2 E70356	chromosome assembl
45	109	10.5	278	1 TPHTWT	troponin T, slow s

ALIGNMENTS

RESULT 1

JC7830
cytokine-inducible 29K protein, CIP29 - human
C:Species: Homo sapiens (man)
C:Date: 09-Dec-2002 #sequence_revision 09-Dec-2002 #text_change 09-Jul-2004
C:Accession: JC7830; PC7188
R:Fukuda, S.; Wu, D.W.; Stark, K.; Pelus, L.M.
Biochem. Biophys. Res. Commun. 292, 593-600, 2002
A:Title: Cloning and characterization of a proliferation-associated cytokine-inducible p
A:Reference number: JC7830; MUID:21920340; PMID:119222608
A:Accession: JC7830
A:Molecule type: DNA
A:Residues: 1-210 <FUK>
A:Cross-references: UNIPROT:P82979; GB:bankit447886
A:Accession: PC7188
A:Molecule type: protein
A:Residues: 121-137;152-168 <FU2>
C:Comment: This protein that is a novel erythropoietin-induced protein is involved in no
either phosphorylated or glycosylated, and is involved in DNA transcription. This prote
C:Genetics:
A:Gene: cip29
A:Map position: 7; 8; 12

Query Match		100.0%;	Score 1040;	DB 2;	Length 210;
Best Local Similarity		100.0%;	Pred. No. 4.4e-57;		
Matches 210;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MATETVELHLKLAELKQEC	LARGLETGKIQD	LIHRLQAYLEHAE	EENEEDVLGDET 60
Db	1	MATETVELHLKLAELKQEC	LARGLETGKIQD	LIHRLQAYLEHAE	EENEEDVLGDET 60
Qy	61	EEETKPIELPVKEEPEKTV	DAAEKVKVITSET	POTERMOKRAERFNP	VSLSKK 120
Db	61	EEETKPIELPVKEEPEKTV	DAAEKVKVITSET	POTERMOKRAERFNP	VSLSKK 120
Qy	121	AAARAARFGISSVPTKGL	SSDNKPMVNDLKL	KERAQRFGNLVSSIS	SRKSEDDKLKKRER 180
Db	121	AAARAARFGISSVPTKGL	SSDNKPMVNDLKL	KERAQRFGNLVSSIS	SRKSEDDKLKKRER 180
Qy	181	FGIVTSSAGTGTTEDTEA	KKRAERFGIA 210		
Db	181	FGIVTSSAGTGTTEDTEA	KKRAERFGIA 210		

RESULT 2

T41291
conserved hypothetical protein SPOC3H12.03c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T41291
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, September 1998

A;Reference number: Z21985
A;Accession: T41291
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-245 <WOO>
A;Cross-references: UNIPROT:O74871; EMBL:AL031824; PIDN:CAA21220.1; GSPDB:GN00068; SPDB:
A;Experimental source: strain 972h-; cosmid c31H12
C;Genetics:
A;Gene: SPDB:SPC31H12.03C
A;Map position: 3

Query Match 21.3%; Score 222; DB 2; Length 245;
Best Local Similarity 29.4%; Pred. No. 5.9e-07;
Matches 74; Conservative 30; Mismatches 80; Indels 68; Gaps 8;

QY 11 LKIAELKQECIARGLTKG1KQDILHRLQAYLEHAEAEANEE----- 53
DB 7 LKVAELREKLAELKGLTAGNKAELVSLTAATESNDENTSNNAATGLDGLAPEDDIDWG 66
QY 54 DVLGDETEEBETKPIELPVKEEPPPEKTVDAEKK--VVKITSEIPQT----- 100
DB 67 DMENDTISTDVKNPAPESKETSAPAAAVEIEKENESIISKTSQAPETSTGAEEHQETT 126
QY 101 -----ERMOKRAERNFNPVSL--KKAARAAFGISSVPTKGLSS 139
DB 127 EESKQSVSNVSSPDVAKEQEKLQRAKRGIPVDDQIKKAAARAFGIO-----QPLAS 182
QY 140 DNKPMVNDLKKERAQFGLNVSISRSKSEDEKLLKKRERFGIVTSSAGTGTTEDETEAK 199
DB 183 SNNKNHNSQNPQRS-----NSRSQRKNKNAPPSAKSKSNIL-----DDPIEA 230
QY 200 K-RKRAERFGIA 210
DB 231 KARKRAERFGVA 242

RESULT 3
T48827
hypothetical protein 68B2.110 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 09-Jul-2004
C;Accession: T48827
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24541
A;Accession: T48827
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-344 <SCH>
A;Cross-references: UNIPROT:Q9F6W5; EMBL:AL353821; GSPDB:GN00112; NCSP:68B2.110
A;Experimental source: cosmid contig 68B2; strain 74
C;Genetics:
A;Gene: NCSP:68B2.110
A;Map position: 2
A;Introns: 50/1

Query Match 14.2%; Score 148; DB 2; Length 344;
Best Local Similarity 27.2%; Pred. No. 0.029;
Matches 62; Conservative 27; Mismatches 95; Indels 44; Gaps 8;

QY 7 ELHKLKLAEKQECIARGLTKG1KQDILHRLQAYLEHAEAEANEDVL-----GDE 59
DB 3 DYSNMKVPELKKLNLNERSLPOTGNKADLIARLQ-----EHDQQAQKPADAAPATTKKDGEA 58
QY 60 TEEBETKPIELPVKEEPPPEKTVDAEKKVVKITSEIPQTERMQKRAERNFNPVSLSK 119
DB 59 EDEIDYEDDDFPAGDKKAATETTD--EKAPASAAAEATATESEPK-----APVESEAQ 109
QY 120 KAARAARFGISSVPTKGLSSDNKPMVNL-----DKLKERAQR-----FGLNVS 162
DB 110 KTTAAPAAPAAASTQPAASQAQGEVAAKDTTTATATTAEEDPPSKAEQKPAEPLFSQHL 169
QY 163 SISRKSEDEKLLKKRERFGIVTSSAGTGTTEDETEAKRKRERFGIA 210

DB 170 PTDAKSE-AEKRAARAARFGITTD-----KSEEAQKARAARFGIA 210

RESULT 4
T48298
hypothetical protein F9G14.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C;Accession: T48298
R;Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24491
A;Accession: T48298
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-214 <BEV>
A;Cross-references: UNIPROT:O9LZ08; EMBL:AL162973
A;Experimental source: cultivar Columbia; BAC clone F9G14
C;Genetics:
A;Map position: 5
A;Introns: 109/2; 138/2; 184/2; 205/3
A;Note: F9G14.80

Query Match 13.4%; Score 139; DB 2; Length 214;
Best Local Similarity 31.2%; Pred. No. 0.062;
Matches 55; Conservative 22; Mismatches 67; Indels 32; Gaps 7;

QY 57 GDETEEETKPIELPVKEEPPPEKTVD-----VAAEKVVKITSEIPQT----- 100
DB 17 GLSTGENPKIVDLNLTTELDRTDILDCVKGFSGEGKEETDSNGICSTAGVDSGLS 76
QY 101 -----ERMOKRAERNFNPVSL--ESKKAARAARFGISSVPTKGLSSDNKPMVNDLKKER 153
DB 77 PVDDIQKIRRAERFGVSVKLTTEEKRNSRAERFGTVAAAVVNGSGTGKKAEL-KRKAR 135
QY 154 AQRFGI-LNVSSISRSKSEDEKLLKKRERFGIVTSSAGTGTTEDETEAKK-K-RAERF 207
DB 136 ADRFGVFSATSTDTKTEEAKKKARLAFRGKTK-----VDSAEENKKRKAALRF 185

RESULT 5
T22490
hypothetical protein Y37A1B.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T22490; T26618
R;Matthews, L.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19569
A;Accession: T22490
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1222 <WIL>
A;Cross-references: UNIPROT:Q9XTH8; EMBL:Z822268; PIDN:CAB05201.1; GSPDB:GN00022; CESP:Y37A1B.1
A;Experimental source: clone F52B11
R;McMurray, A.
submitted to the EMBL Data Library, June 1998
A;Reference number: Z20245
A;Accession: T26618
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1222 <W12>
A;Cross-references: EMBL:AL023835; PIDN:CAA19496.1; GSPDB:GN00022; CESP:Y37A1B.1
A;Experimental source: clone Y37A1B
C;Genetics:
A;Gene: CESP:Y37A1B.1
A;Map position: 4
A;Introns: 78/3; 127/2; 165/3; 450/1; 500/2; 620/3; 728/3; 1080/1; 1157/3

Query Match 13.4%; Score 139; DB 2; Length 1222;
Best Local Similarity 27.3%; Pred. No. 0.38;

Matches	60;	Conservative	34;	Mismatches	82;	Indels	44;	Gaps	9
Qy	11	LKLAELKQECIARGLETGKIQRDILHRLQAYLEEHAEHEEANEEDVLGDTEEEETKPIEL	70						
Db	663	MKAELRVELRGLETGKITLLVQLTALD--TEKAAESAQAARDVMERDA--AEN	717						
Qy	71	PVKSEEPPEK-----TVDAARKVVKITSETIPQTERMOKRAERNFVPVSLESKKAAR	123						
Db	718	AVKOGGEENPAAFIAPISIEETAKTAEAAKKEAEAKRKKEBEULE-----KEKKKR	772						
Qy	124	AARFGISSVPTKGLSSDNKPWNLDKLKRAORFG---LNVSSISRKSSEDDEK-----	173						
Db	773	EALEKHVLP-----KDKKILVFPSK-SFKSGKGCDCKVLSSILLDYRDDNKENQNEFS	826						
Qy	174	-----LKRRKERFGIVTSSAGTGTTEDTEAKKRRAE	205						
Db	827	LFABAFKEMIERNAAFTIVETLANGC---DRDAEKGRDE	863						
RESULT 6									
T34418									
hypothetical protein F12P3.3 - Caenorhabditis elegans									
C:Species: Caenorhabditis elegans									
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999									
C:Accession: T34418									
F:Fulton, B.; Wohldmann, P.									
submitted to the EMBL Data Library, July 1998									
A:Description: The sequence of C. elegans cosmid F12P3.									
A:Reference number: Z21521									
A:Accession: T34418									
A>Status: preliminary; translated from GB/EMBL/DDBJ									
A:Molecule type: DNA									
A:Residues: 1-3488 <FUL>									
A:Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12P3.3									
A:Experimental source: strain Bristol N2; clone F12F3									
C:Genetics:									
A:Gene: CESP:F12P3.3									
A:Map position: 5									
A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1									
Query Match 13.0%; Score 135.5; DB 2; Length 3488;									
Best Local Similarity 23.6%; Pred.No. 1.9;									
Matches 56; Conservative 43; Mismatches 77; Indels 61; Gaps 8;									
Qy	2	ATETVELHKLAELKQCLARGLETGKIQRDILHRLQAYLEEHAEHEEANEEDVLGDTE	61						
Db	1361	ATKADAKLKLKLEQQKAAEIALIEIQEKLAQEQSRIEDEAKSA-EKQKLESETK	1419						
Qy	62	BEETKPIELPVKEBPPEKTVDVAEEKVKITSEIPTERMOK-----BAERNV	112						
Db	1420	SKQT-----BEAPKESVDEKPKKVKKTKSEKSDSSISQKSAKSTVDAETLES	1470						
Qy	113	PVSLESKKAARAARFGISSVPTKGLSSDNKPWNLDKLKRAORPLNVISSIRKSEDE	172						
Db	1471	DFNLVEKKTVOQ----VEQPDSDESTSATIK-----RDPAQK----TEBISKDDGDE	1514						
Qy	173	K-----LKRRKERFGIVTSSAG-----TGTTEDTEAKKRK	203						
Db	1515	KTTTTDGKPPPEDSEATPKGVVKKTKQKSDSVASDASLADVSKLSDDVKEKPKK	1571						
RESULT 7									
T05005									
hypothetical protein T19Pi9.70 - Arabidopsis thaliana									
C:Species: Arabidopsis thaliana (mouse-ear cress)									
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004									
C:Accession: T05005									
R:Bevan, M.; Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Hoheisel, J.; Mewse,									
submitted to the Protein Sequence Database, April 1998									
A:Reference number: Z15394									
A:Accession: T05005									
A:Molecule type: DNA									
A:Residues: 1-633 <BEV>									

```
C;Accession: I51116
R;Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A;Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re
A;Reference number: I51116; MUID:95287814; PMID:7770000
A;Accession: I51116
A;Status: preliminary; translated from GB/EMBL/DBJB
A;Molecule type: mRNA
A;Residues: 1-1110 <JAC>
A;Cross-references: UNIPROT:Q91255; EMBL:U19361; NID:G632548; PIDN:AAA80106.1; PID:G632529
C;Superfamily: neurofilament triplet H protein

Query Match      12.0%; Score 125; DB 2; Length 1110;
Best Local Similarity 24.1%; Pred. No. 2.5;
Matches 46; Conservative 34; Mismatches 85; Indels 26; Gaps 5;

Qy    30 IKQDLIRLOAYLEHAEHEEANEEDVLGDTEEBETKPIELPVKEEPPEPTVDVAEEKK 89
Db        :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
660 VEEREEAAEAEEAEAEAEAEAGEEDVEAESKEEEEEDSKEADAEDEAAEEEV---KEEE 716

Qy    90 VVKITSIPOTERMOKKAERNPNVPVLSKK-----KAAPARFGISSVPTKGLSS 139
Db        :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
717 VTKSDAEEAEAE-AEEESAKSEEAEEAEKDEAEFEAEAEAEVEETEATAEAEAKEASD 775

Qy   140 DNKPMMVNLDKLKERAOFRGLNVSSIISKSDDEKLKKRKRFIVTSSAGTGTTEDTEAK 199
Db        :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
776 DEKP-----EEVKSEAPVPAEKAKPEPKAAPKK-----APAKVESPISEPEDEP 823

Qy    200 KRKRAERFGIA 210
Db      824 KAEVVEKKGA 834

RESULT 10
S00030
neurofilament triplet M protein - mouse
N;Alternate names: 160K NF protein; NF-M
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 09-Jul-2004
C;Accession: S00030
R;Levy, E.; Liem, R.K.H.; d'Eustachio, P.; Cowan, N.J.
Eur. J. Biochem. 166, 71-77, 1987
A;Title: Structure and evolutionary origin of the gene encoding mouse NF-M, the middle-m
A;Reference number: S00030; MUID:87246694; PMID:3036526
A;Accession: S00030
A;Molecule type: DNA
A;Residues: 1-849 <LEV>
A;Cross-references: UNIPROT:P08553; GB:X05640; NID:G53357; PIDN:CAA29127.1; PID:g297529
C;Genetics:
A;Map position: 14
A;Introns: 359/3; 401/2
C;Superfamily: cytoskeletal keratin
C;Keywords: coiled coil; intermediate filament
F;99-244/Domain: coiled coil <COI>
F;267-411/Domain: coiled coil <CO2>
F;412-849/Domain: tail <TAI>

Query Match      11.9%; Score 123.5; DB 2; Length 849;
Best Local Similarity 23.6%; Pred. No. 2.3;
Matches 55; Conservative 30; Mismatches 57; Indels 91; Gaps 11;

Qy    43 EEHAEEANBEDVLGDTEEBETKPI-----ELPVKEE---EPPEKTVD----- 83
Db      EEEGETEAEGE--GEEAAKEEKIKGVVEAVAKGEIKVKEPKAKSPMPKSPVVEVK 617

Qy    84 -----VAAEKKVKVTSETPOTERMOKKAERNPNVPVLSLESKCAAPAAR 126
Db      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
618 KPPEAKAGQEQQKEEKVEEEK---EVTKGSPKEEKVEEKEEK---PKVDADKKAE--- 669

Qy   127 FGISSVPTKGL-----SSDNKPMVNLDKLERAOFRG----- 158
Db      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
670 ---SPVKEKAVEEVIITSKSVKVSLEKDTKEEKFPQP-EKVKEKAEBEGGSEBGSRSRP 725
```

A:Accession: T21174
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2020 <WIL>
A:Cross-references: UNIPROT:Q19658; EMBL:Z75538; PIDN:CAA99841.1; GSPDB:GN00019; CESP:F20G4
A:Experimental source: clone F20G4
R:McMurray, A.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19982
A:Accession: T25113
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2020 <WIL2>
A:Cross-references: EMBL:Z75550; PIDN:CAA99931.1; GSPDB:GN00019; CESP:F20G4.3
A:Experimental source: clone T22C1
C:Genetics:
A:Gene: CESP:F20G4.3
A:Map position: 1
A:Map position: 133/3; 155/2; 195/2; 270/3; 325/1; 1023/3; 1072/2; 1125/2; 1207/3; 1248/3; 13
C:Introns: 133/3; 155/2; 195/2; 270/3; 325/1; 1023/3; 1072/2; 1125/2; 1207/3; 1248/3; 13
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:106-798/Domain: myosin motor domain homology <MWO>

Query Match 11.6%; Score 120.5; DB 2; Length 2020;
Best Local Similarity 24.8%; Pred. No. 8.8;
Matches 56; Conservative 32; Mismatches 91; Indels 47; Gaps 8;

Qy 7 ELHLKLAELKQECIARGLGTGKIGKQDLIHRLOAYLEEHAEAEEDVL-----GDE 59
Db 1153 ETNKEAARQAE-----KARRDMAEELESYKQE--LBSNDKTVLHSQLKAKRDE 1201
Qy 60 TEEETKPIELPVKEEPEPEKTVDAAEKKVVKITSEIPQTERMOK----- 105
Db 1202 EYAHLOKQLEETVKSSEVVEEMKAQNKIBELNETIDQLKR-QKISADKAKSAESDN 1260
Qy 106 ---RAERFNV-PVSLSEKKAARAARFGISSVPTKGLSSDNK---PMVNLDKLERQRF 158
Db 1261 ENFRAELNSIASARLEAEKKRKAEB-----TSLMEKHDKKREMQSNLDDLWAKLSKMN 1313
Qy 159 LNVSSIIRKSEDDKELKKRKEFGIVTSAGTGTEDTEAKKRKA 204
Db 1314 NELESIOKAKSADETLNSNLLKNASLDQLSELTEASEEDRRTRA 1359

RESULT 15
gelsolin-related protein GRP125 - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30330
R:Stocker, S.; Hiery, M.; Marriott, G.
Mol. Biol. Cell 10, 161-178, 1999
A:Title: Phototactic migration of Dictyostelium cells is linked to a new type of gelsolin
A:Reference number: Z20823; MUID:99096692; PMID:9880334
A:Accession: T30330
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1087 <STO>
A:Cross-references: UNIPROT:O96923; EMBL:U95159; NID:g4100185; PID:g4100186; PIDN:AAD007
C:Genetics:
A:Introns: 137/1

Query Match 11.5%; Score 119.5; DB 2; Length 1087;
Best Local Similarity 24.1%; Pred. No. 5.3;
Matches 56; Conservative 38; Mismatches 103; Indels 35; Gaps 9;

Qy 4 ETVELHKLKLAELKQECIARGLGTGKIGKQDLIHRLOAYLEEHAEAEEDVLGDETEEE 63
Db 724 EKEDLEKLK-QQEQEQEQQQQKNNKIVVEVKVEVK---EEDYKVEEVEEVEVKKE 779
Qy 64 ETKPIELPVKEEPEPEKTVDAAB-KKVVKITSEIPQTE-----RMQKRAERFNVPSLE 117
Db 780 EVKSEVAKETKEETKEEYVNDSEATEVKEVQNQVVEEVEEVEEVEEVEEVEEVEEVEE 839

```
Qy 118 SKAARAARFGISSVPTKGLSSDNKPMVNLDKLKERAZQRFGLNVSSISRKSED---DEKL 174
Db 840 EVKEEVKEEVKE-EVKEVKEEVKEEVKQKKEEVNEEIKKEETKEETKEDDNKEDKV 898
Qy 175 KKRKE-----RFGIVTS-----SAGTGTGTEDTEAK-----KKRAE 205
Db 899 NEENETVNEENEVGIIVSPPEKVDDEANSSSTISSPENEGSVSVKDKRKSNE 950
```

Search completed: October 5, 2005, 05:59:36
Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 5, 2005, 00:10:27 ; Search time 87 Seconds
(without alignments)
1236.053 Million cell updates/sec

Title: US-09-788-476A-2
Perfect score: 1040
Sequence: 1 MATETVELHKLALBKQEC.....GTTTETAKKRAERFGIA 210

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1035	99.5	209	1 HCC1 HUMAN	P82979 homo sapien
2	990	95.2	209	1 HCC1 MOUSE	Q9dlj3 mus musculu
3	330	31.7	756	2 Q6Y4V2	Q6Y4V2 bovine vira
4	276.5	26.6	363	2 Q7QB08	Q7QB08 anopheles g
5	222	21.3	245	1 YCM3 SCHPO	O74871 schizosacch
6	222	21.3	253	2 Q9N3G0	Q9N3G0 caenorhabdi
7	222	21.3	308	2 Q9VHC8	Q9VHC8 drosophila
8	165.5	15.9	209	2 Q967H3	Q967H3 hydra atten
9	155.5	15.0	246	2 Q6C2C7	Q6C2C7 yarrowia li
10	155	14.9	1271	2 Q9V8K6	Q9V8K6 drosophila
11	151.5	14.6	203	2 Q6DE01	Q6deq1 xenopus tro
12	148	14.2	344	2 Q9P6W5	Q9P6W5 neurospora
13	144	13.8	784	2 Q6NUT5	Q6nut5 brachydanio
14	143	13.8	177	2 Q8H8U3	Q8h8u3 oryza sativ
15	139	13.4	214	2 Q9LZ08	Q9LZ08 arabidopsis
16	139	13.4	305	2 Q9L327	Q9L327 plasmodium
17	139	13.4	1222	2 Q9XTH8	Q9xth8 caenorhabdi
18	137.5	13.2	204	2 Q8BIQ7	Q8biq7 mus musculu
19	137.5	13.2	1150	2 Q6X935	Q6x935 homo sapien
20	136.5	13.1	230	2 Q6FPLO	Q6fpl0 candida gla
21	136.5	13.1	718	2 Q9NV00	Q9nvq0 homo sapien
22	136.5	13.1	1150	2 Q8IX12	Q8ix12 homo sapien
23	135.5	13.0	329	2 Q6GQC9	Q6gqc9 xenopus lae
24	135.5	13.0	18519	2 Q8ISR6	Q8isf6 caenorhabdi
25	135.5	13.0	18534	2 Q8ISF7	Q8isf7 caenorhabdi
26	135	13.0	774	2 Q9VHY5	Q9vhy5 xenopus lae
27	134.5	12.9	10578	2 Q8ISF5	Q8isf5 caenorhabdi
28	133.5	12.8	315	2 Q8GL99	Q8gl99 streptococc
29	133.5	12.8	317	2 Q8GL86	Q8gl86 streptococc
30	133.5	12.8	633	2 Q6S655	Q6s655 arabidopsis
31	133	12.8	843	2 Q6AXC9	Q6axc9 mus musculu

32	133	12.8	1146	2	Q8CH18	Q8ch18 mus musculu
33	131	12.6	719	2	Q7S8Q9	Q7s8q9 neurospora
34	130.5	12.5	314	2	Q8GLB0	Q8glb0 streptococc
35	130.5	12.5	402	2	Q8U3W6	Q8u3w6 pyrococcus
36	129	12.4	539	2	Q8BJY0	Q8bjy0 m mus muscu
37	128	12.3	1157	2	Q641G3	Q641g3 xenopus lae
38	128	12.3	2006	2	Q7KSQ6	Q7ksq6 plasmodium
39	128	12.3	2019	2	Q7KSQ5	Q7ksq5 plasmodium
40	128	12.3	2055	2	Q8IHP3	Q8ihp3 plasmodium
41	128	12.3	2055	2	Q8TSC7	Q8tsc7 plasmodium
42	127.5	12.3	804	2	Q6PIZ0	Q6piz0 homo sapien
43	126.5	12.2	326	2	Q8GL91	Q8gl91 streptococc
44	126.5	12.2	580	2	P91156	P91156 caenorhabdi
45	126.5	12.2	809	2	Q6PAR2	Q6par2 mus musculu

ALIGNMENTS

RESULT 1
ID HCC1 HUMAN STANDARD; PRT; 209 AA.
AC P82979; Q9P066;
DT 16-OCT-2001 (Rel. 40, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Nuclear protein Hcc-1 (HSPC316) (Proliferation associated cytokine-
DE inducible protein CIP29).
GN Name=HCC1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.; SEQUENCE OF 109-118; 156-166 AND 180-198, TISSUE
SPECIFICITY, AND SUBCELLULAR LOCATION.
RC TISSUE=Liver;
RX MEDLINE=21255661; PubMed=11356193; DOI=10.1016/S0014-5793(01)02409-7;
RA Choong M.L., Tan L.K., Lo S.L., Ren E.-C., Ou K.L., Ong S.-E.,
RA Liang R.C.M.Y., Seow T.K., Chung M.C.M.;
RT "An integrated approach in the discovery and characterization of a
RT novel nuclear-protein-over-expressed in liver and pancreatic tumors."
RL FBS Lett. 496:109-116(2001).
RN [2]
SEQUENCE FROM N.A.
RP TISSUE=Umbilical cord blood;
RC MEDLINE=20499367; PubMed=11042152; DOI=10.1101/gr.140200;
RX Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,
RA Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,
RA Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.;
RT "Cloning and functional analysis of cDNAs with open reading frames for
RT 300 previously undefined genes expressed in CD34+ hematopoietic
stem/progenitor cells."
RL Genome Res. 10:1546-1560(2000).
RN [3]
SEQUENCE FROM N.A.
RP TISSUE=Breast cancer;
RC MEDLINE=21920340; PubMed=11922608; DOI=10.1006/bbrc.2002.6680;
RX Fukuda S., Wu D.W., Stark K., Pelus L.M.;
RT "Cloning and characterization of a proliferation-associated cytokine-
RT inducible protein, CIP29."
RL Biochem. Biophys. Res. Commun. 292:593-600(2002).
RN [4]
SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Schmutz J.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RN SEQUENCE OF 1-9 AND 126-134, AND ACETYLATION SITE ALA-1.
RN Blenvnut W.V., Potts A., Brablan J., Quadroni M.;
RA Unpublished observations (JUI-2004).
CC -1- FUNCTION: May have nucleic acid binding capability that may
CC participate in important transcriptional or translational control
CC of cell growth, metabolism and carcinogenesis.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Low expression in spleen, liver, pancreas,
CC testis, thymus, heart, and kidney. Increased levels are seen in
CC hepatocellular carcinoma and pancreatic adenocarcinoma.
CC -1- INDUCTION: By erythropoietin.
CC -1- SIMILARITY: Contains 1 SAP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; A409089; CAC37950.1; -;
DR ENBL; AF161434; AAF28994.1; -;
DR ENBL; AF486281; AAM09686.1; -;
DR ENBL; BC007099; AAH07099.1; -;
DR PR; JC7830; JC7830.
DR HtAct; P82979; -;
DR H-InvDB; HIX0010706; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003676; F:nucleic acid binding; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR GO; GO:0006445; P:regulation of translation; NAS.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR PROSITE; PS50800; SAP; 1.
KW Acetylation; Direct protein sequencing; DNA-binding; Nuclear protein;
KW Phosphorylation; Transcription; Transcription regulation;
KW Translation regulation.
FT INIT MET 0 0
FT MOD_RES 1 1 N-acetylalanine.
FT DOMAIN 7 41 SAP.
FT MOD_RES 162 162 Phosphoserine (By similarity).
FT CONFLICT 126 126 F -> V (in Ref. 2).
FT CONFLICT 133 148 TGLSSDNKEWNLDK -> NKRSVILITNLWLTWIS (in
FT Ref. 2).
FT CONFLICT 152 156 RAQRF -> ELKDL (in Ref. 2).
FT CONFLICT 198 209 KKKRAERFGIA -> RRGKEQALGPDPKFLILSVLQCF
FT PFLSFFLVTPYKCTVMCLRPASQ (in Ref. 2).
FT SEQUENCE 209 AA; 23539 MW; F24C5B1CF67C073A CRC64;
Query Match 99.5%; Score 1035; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. No. 3.1e-51;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ATETVELHKLKLAELKQECILARGLGTGKIQDILHRLQVLEHAEENEEDVLGDTE 61
Db 1 ATETVELHKLKLAELKQECILARGLGTGKIQDILHRLQVLEHAEENEEDVLGDTE 60
QY 62 EEBTKPIELPVKEEPPKTDVAAEKVKVITSEIPQTERMQKRAERFNVPSVLESKKA 121

Db 61 EEBTKPIELPVKEEPPKTDVAAEKVKVITSEIPQTERMQKRAERFNVPSVLESKKA 120
QY 122 ARAARFGISVPTKGLSSDNKPMVNLDKLKERQRFGLNVSSISRKSEDEKLUKKRKRF 181
Db 121 ARAARFGISVPTKGLSSDNKPMVNLDKLKERQRFGLNVSSISRKSEDEKLUKKRKRF 180
QY 182 GIVTSSAGTGTTEDEAKKRAERFGIA 210
Db 181 GIVTSSAGTGTTEDEAKKRAERFGIA 209
RESULT 2
HCC1_MOUSE STANDARD; PRT; 209 AA.
ID HCC1_MOUSE
AC Q9DIJ3; Q9CU18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Nuclear protein Hcc-1.
DE NamesHcc1;
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J, and NOD; TISSUE=Cerebellum, Embryo, and Thymus;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RX Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schraml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gusticich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Kongsaya A., Kurochkin I.V., Lee Y., Ienhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RT Nature 420:563-573(2002).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman A.C., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A., Smailus D.E.,
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP PHOSPHORYLATION SITE SER-162.
RX PubMed-14729942; DOI=10.1074/mcp.D300003-MCP200;
RA Shu H., Chen S., Bi Q., Mumby M., Brekken D.L.;
RT "Identification of phosphoproteins and their phosphorylation sites in
RT the WHI-231 B lymphoma cell line";
RL Mol. Cell. Proteomics 3:279-286(2004).
CC -!- FUNCTION: May have nucleic acid binding capability that may
CC participate in important transcriptional or translational control
CC of cell growth, metabolism and carcinogenesis (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 1 SAP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK003453; BAB22799.1; -
DR EMBL; AK003453; BAB31400.1; -
DR EMBL; AK088266; BAC40246.1; -
DR EMBL; BC027510; AAH27510.1; -
DR MGD; MGI:1913368; 1110005A23rik.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS00800; SAP; 1.
KW Acetylation; DNA-binding; Nuclear protein; Phosphorylation;
KW Transcription; Transcription regulation; Translation regulation.
FT INIT MET 0 0 By similarity.
FT MOD RES 1 1 N-acetylalanine (By similarity).
FT DOMAIN 7 41 SAP.
FT MOD RES 162 162 Phosphoserine.
FT MOD RES 209 AA; 23401 MW; 719300BAE5C76B7 CRC64;
SQ SEQUENCE 209 AA; 95.2%; Score 990; DB 1; Length 209;
Query Match 95.2%; Score 990; DB 1; Length 209;
Best Local Similarity 95.7%; Pred. No. 1.1e-46;
Matches 200; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ATETVELHLKLAELKQELCLARGLETGKIQDLIHLRQAYLEHAEENEEDVLGDETE 61
Db 1 AAETVELHLKLAELKQELCLARGLETGKIQDLIHLRQAYLEHAEENEEDVLGDETE 60

Qy 62 EETKPTLPVKEEPEKTVDAEAKVVKITSEIPQTERMQRAERFNPVVSLESKKA 121
Db 61 EEPKPTLPVKEEPEKTVDAEASEKVKVITSGIPQTERMQRAERFNPVVSLESKKA 120

Qy 122 ARAARFGISSVPTKGLSSDNKPMWNLDKLKERARQFGLNVSSISRKSEDDKLRKRF 181
Db 121 ARAARFGISSVPTKGLSSDTKPMWNLDKLKERARQFGLNVSSISRKSEDDKLRKRF 180

Qy 182 GIVTSSAGTGTTEDTEAKRRAERFGIA 210
Db 181 GIVTSSAGTGTTEDTEAKRRAERFGIA 209

RESULT 3
Q6Y4V2 PRELIMINARY; PRT; 756 AA.
AC Q6Y4V2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Polyprotein (Fragment).
OS Bovine viral diarrhea virus (BVDV) (Mucosal disease virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Pestivirus.
OX NCBI_TaxID=11099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CP8;
RX PubMed=12970452; DOI=10.1128/JVI.77.19.10663-10669.2003;
RA Mueller A., Rinnick G., Thiel H.-J., Tautz N.;
RT "Cell-derived sequences in the N-terminal region of the polyprotein of
RT a cytopathogenic pestivirus";
RL J. Virol. 77:10663-10669(2003).
DR EMBL; AY182136; AAC86641.1; -
DR GO; GO:0003677; F-DNA binding; IEA.
DR GO; GO:0016032; P-viral life cycle; IEA.
DR GO; GO:0019082; P-viral protein processing; IEA.
DR InterPro; IPR008751; Peptidase_C53.
DR InterPro; IPR003034; SAP.
DR Pfam; PF05550; Peptidase_C53; 1.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS00800; SAP; 1.
KW Polyprotein.
FT NON TER 756 756
SQ SEQUENCE 756 AA; 85191 MW; 71D0EE9E05B669D8 CRC64;
Query Match 31.7%; Score 330; DB 2; Length 756;
Best Local Similarity 37.2%; Pred. No. 7.7e-11;
Matches 97; Conservative 24; Mismatches 68; Indels 72; Gaps 8;

Qy 12 KLAELKQELCLARGLETGKIQDLIHLRQAYLEHAEENEEDVLGDETEETKPTLP 71
Db 445 KLAELKQELCLARGLETGKIQDLIHLRQAYLEHAEENEEDVLGDETEETKPTLP 504

Qy 72 VKEEPEPEKTVDA-----AEKKV--KITSEIPQTE 101
Db 505 VKEESARQGTAKLTWLGRLGILKKLEEPVVDQAGNPLFGRGVHPQSTLKL-- 561

Qy 102 RMQKRAERFNPVVSLESKKAARAFGISVPTKGLSSDNKPMWNL-----KL 150
Db 562 --HKRGER-EVPTNLASLPKRGDCRSNDKGPVSGIYKPGPLFYQDYEGPVYHRAPLEL 618

Qy 151 KERA-----QRGLNVSSISR-----KSEDEKLKK---RKRFQIVT 185
Db 619 FEETMCEVTKRIGRVGTGSDSKLYHYVICDGCIIVKSATKDRQKVLKWNKNCPLWV 678

Qy 186 SSAGTGTTEDTEAKRRAER 206
Db 679 SSCSDTKDEGAVRKKQKQKPD 699

RESULT 4
Q7QBC8 PRELIMINARY; PRT; 363 AA.
AC Q7QBC8;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE AgCP2865 (Fragment).
GN Name=agCS5572; ORFNames=ENSANGG00000012227;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein Y53G8AR.6.
GN Name=Y53G8AR.6; ORFNames=Y53G8AR.6;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Latreille P.;
RT "The sequence of C. elegans cosmid Y53G8AR.6";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.

RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilton R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RN [14]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024809; AAF59543.1; -;
DR WormBase; WBGene00021813; Y53G8AR.6.
DR WormPep; Y53G8AR.6; CE25422.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS00800; SAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 253 AA; 26960 MW; F573AD88CD188DA1 CRC64;

Query Match 21.3%; Score 222; DB 2; Length 253;
Best Local Similarity 30.3%; Pred. No. 3e-05;
Matches 79; Conservative 38; Mismatches 80; Indels 64; Gaps 11;

Qy 1 MATETVELHKLKLAELKQECIARGLETGKIQDLIHLQALYLEHAEAEANE----- 52
Db 1 MVLSDADAKKLTVPRLKEELASGLDNTGNKPELLAELTESIQAADELGLDAGITS DTH 60
Qy 53 ---EDVLGDE-----TEETKPI-----ELPVKEEPEPTKVDAAEKKVVKIT 94
Db 61 LLSDDILNDDILDAPSVDGESEADKLLGSVDGAEEHKTDKSEGEKHEVHLIDKKIDEVT 120
Qy 95 SEIPQTERMQKRAERFNVVSVLE-----SKQAARARFGISSVPTKGLSSDNK----- 142
Db 121 AE---TDK-RARAIRFGLPVTAEIDLGS DGAKAARAKRFDLPEDSKRLTSDAKASRAERF 176
Qy 143 -----PMVNLDKLERAOQRFGLNV-----SSISRKSEDDDE----KLKKRKERFGITSS 187
Db 177 GITPKTPAAD-DKLAARARFGIPVGGASASAKNAPNSEGKAAKLAARAKRFG----- 230
Qy 188 AGTGTTTDEAKRKRAERFG 208
Db 231 -GEIDDEMEAKKKARMERFG 250

RESULT 7
Q9VHC8 PRELIMINARY; PRT; 308 AA.
AC Q9VHC8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE CG8149-PA (LD31448p).
GN OREName=CG8149;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abiril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harrie N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jimali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celnik S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celnik S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a Genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Radecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celnik S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RN SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RN SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RN [7]
RN SEQUENCE FROM N.A.
RC STRAIN=Berkeley;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Faragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celnik S.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS003683; AAF54386.1; -;
DR EMBL; AY061392; AAL28940.1; -;
DR FlyBase; F89n0037700; CG8149.
DR GO; GO:0048102; P:autophagic cell death; IEP.
DR GO; GO:0035071; P:salivary gland cell death; IEP.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS0800; SAP; 1.
SQ SEQUENCE 308 AA; 32904 MW; 3E502371396D102E CRC64;
Query Match 21.3%; Score 222; DB 2; Length 308;
Best Local Similarity 27.3%; Pred. No. 3.7e-05;
Matches 83; Conservative 40; Mismatches 77; Indels 104; Gaps 10;
QY 7 ELHLKLAELKQECCLARGLETGKIKODLTHRLQA-----YLEEHAEENEED----- 54
Db 5 DVTMKVADLKRELKRLGLAVNGKTELQDLRLQTLLEGDLSEDSAIADAIDDDVVSFT 64
QY 55 -----VLGDETEEE-----ETKPIEL----- 70
Db 65 DEDEHKLGDNDDELLKSPVSTPTTVAIPDLLAEKTSAPDAAAPTCKIVLKRNNSSQ 124
QY 71 -----PVKEEPP-----EKTVDV-----AAEKKVVKITS 95
Db 125 STGTVASTGTTTSKENEAPAAASDSTGPTTKKHPVVGPKTEGKSGDKKLNLTA 184
QY 96 EIPQTERMQKRAERFNV-PVSLKSKAARAARFGISSVPTKGLSSNKPWNLDK--LKE 152
Db 185 Q-----ERLELRACKFGITPPAVANTATAVAVAINKSSASITANKGNKETEEQKEALKK 240
QY 153 RAQRFGLNVSSIRKSEDEKLKRRKRGIVTSSAGTGTTEA-----KKRKAER 206
Db 241 RAERFGVVPDKAPTSKADRLQKRERFGAGAVSAATTTPTTTESKDAWSEKARARLER 300
QY 207 FGIA 210
Db 301 FKTA 304
RESULT 8
Q967H3 PRELIMINARY; PRT; 209 AA.
AC Q967H3
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative SAP domain-containing protein (Fragment).
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Sukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6087;
RN [1]
RP SEQUENCE FROM N.A.
RA Stover N.A., Steele R.E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY032618; AAK51143.1; -;
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS0800; SAP; 1.
FT NON TER 209 209
SQ SEQUENCE 209 AA; 23630 MW; 787F9A209DCA2E5 CRC64;
Query Match 15.9%; Score 165.5; DB 2; Length 209;


```
Db 115 ENGDDQGFQEGDEEEDEGIPVGLSEBDDAEAA 151
RESULT 12
Q9P6W5 PRELIMINARY; PRT; 344 AA.
AC Q9P6W5
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein 68B2.110.
GN Name=68B2.110;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL353821; CAB88619.1; -.
DR PIR; T48827; T48827.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR003034; SAP.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS00800; SAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 344 AA; 36133 MW; 5510625BACAF16343 CRC64;
Query Match 14.28; Score 148; DB 2; Length 344;
Best Local Similarity 27.26; Pred. No. 0.63; Mismatches 27; Indels 44; Gaps 8;
Matches 62; Conservative 27;
QY 7 ELHKLAELKQELARGLETGKIGKQDLIHLQAYLEHAEHEEEDVL-----GDE 59
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 DYNMKVPKLPKLLNLSLQTNKADLIARLQ-----EHDQQAQKPADAPATTKDGEA 58
QY 60 TEEETKPIELPVKEEPPKPTVDVAEKKVKVITSEIPOTERMOKRAERNVPVLSK 119
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 EDEIDYDDDFPAGDKKAATETD---EKAPASAAAATATSESPK-----APVESEAQ 109
QY 120 KAARAARFGISSVPTKGLSDNKPVMNL-----DKLKERAR-----FGLNVS 162
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 110 KTTAAPAAPAAASTQPAASQAQGEVAAKDTTATATTAEDPSKAEQKPAELFSQHLP 169
| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 163 SISKSDDEKLKKRERFGIVTSSAGTGTTEDEAKKRAERFGIA 210
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 170 PTDKSE-AKRAARAARFGITDE-----KSEEAQAKAARAARFGIA 210
RESULT 13
Q6NUT5 PRELIMINARY; PRT; 784 AA.
AC Q6NUT5
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Zgc:85971.
GN ORPNames=zgc:85971;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
```

```
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Strausberg R.;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC068436; AAH68436.1; -.
DR ZFIN; ZDB-GENE-040426-2432; zgc:85971.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR003034; SAP.
DR InterPro; IPR003877; SPRY_receptor.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00449; SPRY; 1.
DR SMART; SM00513; SAP; 1.
DR PROSITE; PS00800; SAP; 1.
DR PROSITE; PS05080; SAP; 1.
SQ SEQUENCE 784 AA; 89050 MW; 36C003AD2B49F226 CRC64;
Query Match 13.8%; Score 144; DB 2; Length 784;
Best Local Similarity 26.1%; Pred. No. 2.5;
Matches 47; Conservative 30; Mismatches 73; Indels 30; Gaps 3;
QY 5 TVELHKLKLAELKQELARGLETGKIGKQDLIHLQAYLEHAEHEEEDVLG-DETEE 63
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 SLDVKKLKNELKEELQKGLDGLKADLVERLAALAEDEKQTCDERYETGVAAEGND 61
QY 64 ETRKPIELPVKEEPPKPTVD---VAAEKVKVITSEIPOTERMOKRAERNVPVLSK 120
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 EQSDQEQMAGEKESSENGTDFPEVPNDESKIKSVDSQPAIVETVSKEEHEVAPVKLS 121
QY 121 AARAARFGISSVPTKGLSDNKPVMNLDKLKERARFGILNVSSISRKSEDEKLKKR 180
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122 DI-----KTEEQEQRPEASSEREALLESQTECKTEPDR 155
RESULT 14
Q8H8U3 PRELIMINARY; PRT; 177 AA.
AC Q8H8U3
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,
RA Fadrosh D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
```

```

DR  EMBL; AY050475; AAK91488.1; -.
DR  EMBL; AF378893; AAK53696.1; -.
DR  PIR; T48298; T48298.
KW  Hypothetical protein.
SQ  SEQUENCE 214 AA; 2307B981E9297ADE CRC64;
      Query Match      13.4%; Score 139; DB 2; Length 214;
      Best Local Similarity 31.2%; Pred. No. 1.2;
      Matches 55; Conservative 22; Mismatches 67; Indels 32; Gaps 7;
QY  57  GDETEEEETKPIELPKYEEPEKTVD-----VAEEKVKVITSEIPOT----- 100
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db  17  GLSTGNPKKIIVDLNTELDRTDIILDGEVGSDSGEKKETDSNGIGTAGVSDGIS 76
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY  101 -----ERMOKRAERFNVPSVL--ESKKAARAAFGISSVPTKGLSSDNKPMVNLDKLKER 153
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db  77  PVDYDIQKTRRAERFGVSVKVLTEEEKRNSPAERFGTVAADVNGSGEGTKAAEL-KRKAR 135
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY  154  AQRFGL-NVSSISRKSEDEKLLKKRKERFGIVTSSAGTGTTEDTEAKRK-RARERF 207
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db  136  ADRFGVPSATSTTDKTEEEAKKKARLARFGKETK-----VDSAEENKKRKARLARF 185
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

• OM protein - protein search, using sw model

Run on: October 5, 2005, 05:33:55 ; Search time 29 Seconds
(without alignments)
540.562 Million cell updates/sec

Title: US-09-788-476A-2
Perfect score: 1040
Sequence: 1 MATETVELHKLKLAELKQEC.....GTTEDTAKKRAEREGIA 210

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Issued Patents AA.*
- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	742	71.3	149	4	US-09-513-999C-4813
2	625	60.1	139	4	US-09-621-976-7487
3	143.5	13.8	343	4	US-09-270-767-42868
4	125.5	12.1	218	4	US-09-538-092-241
5	120.5	11.6	824	4	US-09-538-092-1242
6	120.5	11.6	1180	4	US-09-543-681A-6436
7	116	11.2	806	1	US-07-980-528-2
8	115	11.1	428	4	US-09-286-981B-7
9	114.5	11.0	253	4	US-09-270-767-41091
10	114.5	11.0	253	4	US-09-270-767-56307
11	114.5	11.0	1507	3	US-08-929-329-5
12	113.5	10.9	1089	4	US-09-949-016-10326
13	113	10.9	413	4	US-09-286-981B-5
14	112	10.8	817	4	US-09-248-796A-20276
15	111	10.7	406	4	US-09-286-981B-18
16	111	10.7	667	4	US-09-248-796A-14492
17	110.5	10.6	375	4	US-09-538-092-1082
18	110.5	10.6	375	4	US-09-949-016-6369
19	110.5	10.6	386	4	US-09-949-016-10070
20	110.5	10.6	1402	4	US-09-248-796A-14503
21	109	10.5	735	4	US-10-164-595-80
22	109	10.5	779	4	US-10-164-595-56
23	109	10.5	784	4	US-10-164-595-79
24	109	10.5	843	4	US-10-164-595-54
25	108.5	10.4	1231	4	US-08-714-741-41
26	108	10.4	588	4	US-08-714-741-42
27	108	10.4	864	4	US-08-714-741-40

28	107.5	10.3	447	4	US-09-710-279-2900	Sequence 2900, Ap
29	107.5	10.3	450	4	US-09-710-279-3226	Sequence 3226, Ap
30	107.5	10.3	2079	4	US-09-949-016-8301	Sequence 8301, Ap
31	107	10.3	412	2	US-08-741-134-2	Sequence 2, Appli
32	107	10.3	440	4	US-09-762-194-2	Sequence 2, Appli
33	107	10.3	440	4	US-09-762-194-4	Sequence 4, Appli
34	106	10.2	477	4	US-09-248-796A-15439	Sequence 15439, A
35	106	10.2	529	4	US-09-248-796A-16357	Sequence 16357, A
36	106	10.2	1702	3	US-08-296-791-5	Sequence 5, Appli
37	106	10.2	1702	4	US-09-839-996-5	Sequence 5, Appli
38	106	10.2	1702	4	US-10-080-505-5	Sequence 5, Appli
39	106	10.2	1702	4	US-10-645-655-5	Sequence 5, Appli
40	106	10.2	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
41	106	10.2	1786	3	US-08-973-462-8	Sequence 8, Appli
42	105.5	10.1	1076	4	US-09-976-594-889	Sequence 889, App
43	105.5	10.1	1857	4	US-09-917-254-91	Sequence 91, Appl
44	105.5	10.1	1972	4	US-09-538-092-1084	Sequence 1084, Ap
45	105.5	10.1	1984	4	US-09-949-016-7111	Sequence 7111, Ap

ALIGNMENTS

RESULT 1

US-09-513-999C-4813
; Sequence 4813, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US/60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 3668
; SOFTWARE: Patent.pm
; SEQ ID NO 4813
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-4813

Query Match	71.3%	Score	742	DB	4	Length	149
Best Local Similarity	100.0%	Pred. No.	5.2e-62	Mismatches	0	Indels	0
Matches	149	Conservative	0	Indels	0	Gaps	0
Qy	1	MATETVELHKLKLAELKQEC	LARGLETGKIGKQDLIHRLOAYLEHAEANEEDVLGDET	60			
Db	1	MATETVELHKLKLAELKQEC	LARGLETGKIGKQDLIHRLOAYLEHAEANEEDVLGDET	60			
Qy	61	EEETKTIELPVKEEPEPEKTDVAAEKVKVITSEIPQTERMQRAERFNPVVSLESKK	120				
Db	61	EEETKTIELPVKEEPEPEKTDVAAEKVKVITSEIPQTERMQRAERFNPVVSLESKK	120				
Qy	121	AARAARFGISVPTKGLSSDNKPMVNLDK	149				
Db	121	AARAARFGISVPTKGLSSDNKPMVNLDK	149				

RESULT 2

US-09-621-976-7487
; Sequence 7487, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2


```

; APPLICANT: Kiledjian, Megerditch
; APPLICANT: Portman, Douglas S.
; TITLE OF INVENTION: METHODS OF PROMOTING INTERMOLECULAR
; INTERACTIONS INVOLVING A NUCLEIC ACID AND MOLECULES
; TITLE OF INVENTION: INTERACTIONS INVOLVING A NUCLEIC ACID AND MOLECULES
; TITLE OF INVENTION: INTERACTIONS INVOLVING A NUCLEIC ACID AND MOLECULES
; TITLE OF INVENTION: INTERACTIONS INVOLVING A NUCLEIC ACID AND MOLECULES
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 5457026ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25; mb/MD
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/980,528
; FILING DATE: 199211120
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Debuca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: UPN-0847
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 806 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-980-528-2

Query Match 11.2%; Score 116; DB 1; Length 806;
Best Local Similarity 23.4%; Pred. No. 0.019;
Matches 64; Conservative 37; Mismatches 82; Indels 90; Gaps 10;

Qy 1 MATETVELHKLKLAELKQECCLARGLSTKGKODLIHRLQAYLEH-----ABEEAN--- 51
Db 1 MSSSPVNVKLVKSELKELKRLSDKGLKALMERLQAALDDEEAGGRPAMEPGNSL 60
Qy 52 -----EEDVLGDTEEEETK-----PIEL-----P 71
Db 61 DLGDSAGRSAGLGLEQEAAGGDEEEEEESEEGISALDQOMELGEENGAAGAASGP 120
Qy 72 VKEEE---PPEKTVDAAEKKVVKITSE-----IPOTERMQKRAERFNV 112
Db 121 MESEEAASEDNGDDQGFQGEDELDEEERGAGDENGCHGEQQPPATQQQPPQQRG-- 178
Qy 113 PVSLESKKAARAARFGISSYPTKGLSDNKKPMVNLKDKERARFGLNVSSISRKSDDE 172
Db 179 ----AAKEAA-----GKSSGPTSLFAVTVP-----PGARQQQAGGDKTEQKGGDKR 225
Qy 173 KLKRRKERFIVTSSAGTGTTEDTEAKKRKRAE 205
Db 226 GVKRRPRE-----DHGRGYFEYIEENKYSRAK 251

RESULT 8
US-09-286-981B-7
; Sequence 7, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

```

```
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-7

Query Match      11.1%; Score 115; DB 4; Length 428;
Best Local Similarity 26.0%; Pred. No. 0.01;
Matches 59; Conservative 28; Mismatches 90; Indels 50; Gaps 10;

QY 2 ATETVELHLKLAELKQEC--LARGLETGKIQDLIHLRLQAYLE-----EHAEEANEED 54
Db 165 ANEPRDEKIQAEAKVESQAEATRLKKIKTDREQAEATRLNIKTDRQAEAEAKVKD 224
QY 55 -----VLGD-----ETEEETKPIELPVKEEPEKTVDVAAEKVKVITSEIPQTE 101
Db 225 EPKRTKRGVLGPATPDKKXENDAKSSDSVGEETLSPS--LKPEKVAEAEKKVEEAK 282
QY 102 RM---QKAEFNVPSVLESKKAARAARFGISSVPTKGLSSDNKPMVNLDKLKERARQFG 158
Db 283 KKAEDQKEEDRNYPTNTYKTLLETAE-----SDVEVK-----KAELELVKEEA---- 327
QY 159 LNVSSISRKSEDEKLRKRFVITSSAGTGTEDTEAKRKRAE 205
Db 328 -----KEPRNEEKVQAKAE---VESQAEATRLNIKTDRKAE 364

RESULT 9
US-09-270-767-41091
; Sequence 41091, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41091
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-41091

Query Match      11.0%; Score 114.5; DB 4; Length 253;
Best Local Similarity 23.8%; Pred. No. 0.0058;
Matches 55; Conservative 35; Mismatches 80; Indels 61; Gaps 9;

QY 4 ETVELHLKKL---AELKQECLEAR-----GLETGK-----IKQDLIHLRLQAYLEEH 45
Db 55 EDIEVTKVAVLDAELFKACGGRTAHKGARHGLKLSGKIARLEQQEREMLEKLQRKLKTT 114
QY 46 ABEANEEDVLGDETEETEETKPIELPVKEEPEKTVDVAAEKVKVITSEIPQTERMQK 105
Db 115 PETEHRVAVESVDSVEQCDDTK-----AKKKKKSIAEEFSEACEVSQ 155
QY 106 RAERFNVPSVLESKKAARAARFGISSVPTKGLSSDNKPMVNLDKLKERARQFGLNYSIS 165
Db 156 LEE-----PIKSKKKKKDKA--RADROAGLGATGELPQDVTDEQVSKKKRKTAE 253

RESULT 10
US-09-270-767-56307
; Sequence 56307, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56307
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-56307

Query Match      11.0%; Score 114.5; DB 4; Length 253;
Best Local Similarity 23.8%; Pred. No. 0.0058;
Matches 55; Conservative 35; Mismatches 80; Indels 61; Gaps 9;

QY 4 ETVELHLKKL---AELKQECLEAR-----GLETGK-----IKQDLIHLRLQAYLEEH 45
Db 55 EDIEVTKVAVLDAELFKACGGRTAHKGARHGLKLSGKIARLEQQEREMLEKLQRKLKTT 114
QY 46 ABEANEEDVLGDETEETEETKPIELPVKEEPEKTVDVAAEKVKVITSEIPQTERMQK 105
Db 115 PETEHRVAVESVDSVEQCDDTK-----AKKKKKSIAEEFSEACEVSQ 155
QY 106 RAERFNVPSVLESKKAARAARFGISSVPTKGLSSDNKPMVNLDKLKERARQFGLNYSIS 165
Db 156 LEE-----PIKSKKKKKDKA--RADROAGLGATGELPQDVTDEQVSKKKRKTAE 253

RESULT 11
US-08-929-329-5
; Sequence 5, Application US/08929329
; Patent No. 6120770
; GENERAL INFORMATION:
; APPLICANT: Adams, John H
; APPLICANT: Dalton, John P
; APPLICANT: Kappe, Stefan
; TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing
; TITLE OF INVENTION: Vaccine Compositions
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 S Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/929,329
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Breen, John P
; REGISTRATION NUMBER: 38,833
; REFERENCE/DOCKET NUMBER: 835910-28685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
```

```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1507 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
US-08-929-329-5

Query Match      11.0%; Score 114.5; DB 3; Length 1507;
Best Local Similarity 24.9%; Pred. No. 0.059;
Matches 52; Conservative 41; Mismatches 85; Indels 31; Gaps 7;

QY  2 ATETVELHLKLAELKQCLARGLETKGKQDLIHLRLOAVLEHAEAEEDVLGDETE 61
Db  1114 AKAEERKKAEEVKAEEERKRIEAKKAEERKRIEAEKAEERKRIEAEKKA 1169

QY  62 EETPKIPLVKEEPPKTVDAAEKKVVKITSEIPQTERMQKRAERFNVPSLESKKA 121
Db  1170 EEEKIIEAAKAEERKRIEAKKAEERKRIEAKKAEERKKAEE--VKKAEAKK 1227

QY  122 ARAARFGISSVPTKGLSSDNKPMVNLKLERAQRFGLNVSSISRSKSEDEK----LKRR 177
Db  1228 AEA---KALKUKESRGGKTTIEAVKKAEEERKRI-----EAKKAEERKRIEAVKKQ 1279

QY  178 KERFGIVTSSAGTGTDTTDAKKEKRAER 206
Db  1280 KKK-----ENVEAEKKAEEER 1295

RESULT 12
US-09-949-016-10326
; Sequence 10326, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10326
; LENGTH: 1089
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10326

Query Match      10.9%; Score 113.5; DB 4; Length 1089;
Best Local Similarity 29.4%; Pred. No. 0.048;
Matches 55; Conservative 32; Mismatches 77; Indels 23; Gaps 9;

QY  2 ATETVELHLKLAELKQCLARGLETKGKQDL--IHLRLOAVLEHAEAE-----ANEE 53
Db  342 ASHTKPEKQLVKAKLOAKLRFERQKQALFKLDNEDGDEFEDEEEETDSEEE 401

QY  54 DVLGDETEETPKIPLVKEEPPKTVDAAEKKVVKITSEIPQTERMQKRAERFNV 113
Db  402 D--GEERKVEKEEELSEEEKEEENQETAEFLSSSERIET-KOEKENDKENND 458

QY  114 VLESKKAARAARFGISSVPTKGLSSDNKPMVNLKLERAQRFGLNVSSISRSKSEDEK 173
Db  114 VLESKKAARAARFGISSVPTKGLSSDNKPMVNLKLERAQRFGLNVSSISRSKSEDEK 173
```

```

Db  459 GSSEIGKAV-----GFLSVP-KSLSSDSTILL-----PKDSSSKMG--YFPTTEKSEKSETDEN 506

QY  174 LKXKXER 180
Db  507 SGKQPSK 513

RESULT 13
US-09-286-981B-5
; Sequence 5, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizenmann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-5

Query Match      10.9%; Score 113; DB 4; Length 413;
Best Local Similarity 22.8%; Pred. No. 0.015;
Matches 53; Conservative 37; Mismatches 78; Indels 64; Gaps 9;

QY  1 MATETVELHLKLAELKQCLARGLETKGKQDLIHLRLOAYLE-----E 44
Db  138 IAEEDVEVKAELELLKEEAQTR-----NEDTINQAKAKAVESKKAEEATLKKEIKTDRK 190

QY  45 HAESEAN-----REDVLGDE-----TEEEETPKIPLVKEEPPKTV 82
Db  191 KAEERKKAEEAEEDVKDKLRRTKRAVGPGEATPFKENDAKSDSSVGEETLPS- 249

QY  83 DVAEEKVVKITSEIPQTERM---OKEAERFNVPS-----LESKKAARAARFGIS 130
Db  250 -LKSGRKVAEAEKKVAEAKKADQKEEDRRNPTNTTKTLDLEIAESDVVKKEAELELV 308

QY  131 SVPTKGLSSDNKPMVNLKLA--ERAQRFGLVSSISRSKSEDEKLEKKE 179
Db  309 KEEAKG--SRNEEKINQAKAEVESKKAEEATLEKTKTDRKKAEERKKAEE 358

RESULT 14
US-09-248-796A-20276
; Sequence 20276, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20276
; LENGTH: 817
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20276

Query Match      10.8%; Score 112; DB 4; Length 817;
```


GenCore version 5.1.6
 Copyright (c) 1993 - 2005 Compugen Ltd.
 OM nucleic - nucleic search, using sw model
 Run on: October 4, 2005, 19:19:52 ; Search time 4044.85 Seconds
 (without alignments)
 10458.088 Million cell updates/sec

Title: US-09-788-476A-3
 Perfect score: 873
 Sequence: 1 tggagtggagggttaacaaga.....tgattatgtgtgtgtaaat 873

Scoring table: IDENTITY NUC
 Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hhg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	873	100.0	1071	6	BD276528 FULL-LENG
2	873	100.0	1071	6	AX048092 Sequence
3	873	100.0	1520	6	AX405619 Sequence
4	873	100.0	1557	9	HSA409089
5	871.4	99.8	923	9	BC007099 Homo sapi
6	690.8	79.1	2553	6	CQ413602
7	647.4	74.2	84113	9	AC073063
8	633.6	72.6	189507	9	AF486281
9	633	72.5	633	9	AC145981
10	622.4	71.3	220280	2	AC145981
11	599.6	68.7	155666	10	AC134830
12	599.6	68.7	159681	10	AC132088
13	595.2	68.2	2553	6	CQ413602
14	586.2	67.1	908	10	BC027510
15	540.4	61.9	148152	2	AC022200 Homo sapi
16	540.4	61.9	149312	2	AC073328 Homo sapi
17	540.4	61.9	153936	9	AF001207 Homo sapi
18	540.4	61.9	166471	2	AC091052 Homo sapi
19	525	60.1	553	6	AR415990 Sequence

20	525	60.1	553	6	AX972824	Sequence
21	525	60.1	553	6	BD111543	BD111543 EST and e
22	469.2	53.7	539	11	BV179620	BV179620 sqm10607
23	466.2	53.4	254288	2	AC126112	AC126112 Rattus no
24	465.6	53.3	471	6	AX884873	AX884873 Sequence
25	465.6	53.3	471	6	BD024483	BD024483 Sequence
26	458.4	52.5	488	6	CQ684732	CQ684732 Sequence
27	455	52.1	470	6	CQ683930	CQ683930 Sequence
28	431	49.4	470	6	CQ695331	CQ695331 Sequence
29	422	48.3	887	5	AJ719689	AJ719689 Gallus ga
30	415.8	47.6	643	5	EX932722	EX932722 Gallus ga
31	386	44.2	426	6	CQ695645	CQ695645 Sequence
32	379	43.4	379	6	CQ698322	CQ698322 Sequence
33	374.2	42.9	405	6	AR427142	AR427142 Sequence
34	374.2	42.9	405	6	AX987836	AX987836 Sequence
35	374.2	42.9	405	6	BD122695	BD122695 EST and e
36	334	38.3	408	6	AX210422	AX210422 Sequence
37	331.6	38.0	383	6	AR427141	AR427141 Sequence
38	331.6	38.0	383	6	AX987835	AX987835 Sequence
39	331.6	38.0	383	6	BD122694	BD122694 EST and e
40	307.2	35.2	797	5	CR760849	CR760849 Xenopus t
41	265.4	30.4	54666	9	AC073487	AC073487 Homo sapi
42	265.4	30.4	105736	2	AC090678	AC090678 Homo sapi
43	258.6	29.6	266	6	BD071528	BD071528 Secreted
44	255.8	29.3	204050	10	AL672298	AL672298 Mouse DNA
45	241	27.6	337	6	AR426739	AR426739 Sequence

ALIGNMENTS

RESULT 1
 BD276528 1071 bp DNA linear PAT 17-JUL-2003
 LOCUS FULL-LENGTH MOLECULES EXPRESSED IN HUMAN TISSUES.

DEFINITION BD276528
 ACCESSION BD276528
 VERSION BD276528.1 GI:33086296

KEYWORDS JP 2002543839-A/31.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Shih,L.L., AuYoung,J.L., Lu,D.A.M., Lal,P., Batra,S., Tang,T.Y.,
 Yang,J., Azimzai,Y., Reddy,R., Henry, Yue and Baughn,M.R.

TITLE FULL-LENGTH MOLECULES EXPRESSED IN HUMAN TISSUES

JOURNAL Patent: JP 2002543839-A 31 24-DEC-2002;

INCYTE GENOMICS INC,Henry YUE,Tom Y TANG,Preeti KAL,Roopa REDDY,
 Sajeev BAIRA,Mariah R BAUGHN,Jumming YANG,falda AZIMZAI, Dyung Aina

M LU, Janice AU-YOUNG,Leo L SHIH

OS Homo sapiens

PN JP 2002543839-A/31

PD 24-DEC-2002

PP 12-MAY-2000 JP 2000618453

PR 14-MAY-1999 US 09/311937,14-MAY-1999 US 09/311940, PR

14-MAY-1999 US 09/311894

PI leo l shih,janice l au-young,dyung aina

m lu,preeti lal,sajeev

PI batra,

PI tom y tang,jumming yang,yalda azimzai,roopa reddy,henry PI

yue,mariah r baughn

CC This description about <220> can't be interpreted CC <220>

CC <221> misc feature

CC <223> Incyte ID No: 4093555CB1

Key Location/Qualifiers.

1. .1071

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

FEATURES

source

ORIGIN

Query Match 100.0%; Score 873; DB 6; Length 1071;

Best Local Similarity 100.0%; Pred. No. 8.2e-205;			
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	TGGAGT	GAGGGGTAAACAAGATGGCGACCGGCGAGACGGTGAGCTCCATAAGCTAAAGCTTGC 60
Db	165	TGGAGT	GAGGGGTAAACAAGATGGCGACCGGCGAGACGGTGAGCTCCATAAGCTTAAGCTTGC 224
Qy	61	CGAACT	AAAGCAAGAATCTCTTGCTCGTGGTTTGAGACCAAGGGAAATAAGCAAGATCT 120
Db	225	CGAACT	AAAGCAAGAATCTCTTGCTCGTGGTTTGAGACCAAGGGAAATAAGCAAGATCT 284
Qy	121	TATCCAC	AGCTCCAGGCATATCTTGAGAACATCTCTGAAGGAGCGCAATGAGAGAGA 180
Db	285	TATCCAC	AGCTCCAGGCATATCTTGAGAACATCTCTGAAGGAGCGCAATGAGAGAGA 344
Qy	181	TGTACT	TGGGAGTGAACACAGAGGAAGAAACAAAGCCCATTGAGCTCCCTGTCAAAGA 244
Db	345	TGTACT	TGGGAGTGAACACAGAGGAAGAAACAAAGCCCATTGAGCTCCCTGTCNAGA 404
Qy	241	GGAGAA	CCCCCTGAAAAAATGTTGATGTGGCAGCAGAGAAAGATGTTGAAATATAC 300
Db	405	GGAGAA	CCCCCTGAAAAAATGTTGATGTGGCAGCAGAGAAAGATGTTGAAATATAC 464
Qy	301	ATCTGA	ATACACAGACTGAGAGAAATGTCAGAGAGGGCTCAACGATTCATATGATACCTGT 360
Db	465	ATCTGA	ATACACAGACTGAGAGAAATGTCAGAGAGGGCTCAACGATTCATATGATACCTGT 524
Qy	361	GAGCTT	TGGAGATGAAGAAGCTGCTCGGCGAGCTAGTGTGGGATTTCTTCAGTTCACAAC 420
Db	525	GAGCTT	TGGAGATGAAGAAGCTGCTCGGCGAGCTAGTGTGGGATTTCTTCAGTTCACAAC 584
Qy	421	AAAAGGT	CTGTCATCTGATACAAACCTATGGTTAACTTGATTAAGCTGAGAGGAAGAGC 480
Db	585	AAAAGGT	CTGTCATCTGATACAAACCTATGGTTAACTTGATTAAGCTGAGAGGAAGAGC 644
Qy	481	TCAAAG	ATTGGTTGAAATGTCTTCAATCTCCAGAAAGTCTCAAGATGATGAGAAACT 540
Db	645	TCAAAG	ATTGGTTGAAATGTCTTCAATCTCCAGAAAGTCTCAAGATGATGAGAAACT 704
Qy	541	GA AAAA	AGAGAGAGCGAATTTGGATTTGTCACAAGTTCACTGGAACTGGAAACACAGA 600
Db	705	GA AAAA	AGAGAGAGCGAATTTGGATTTGTCACAAGTTCACTGGAACTGGAAACACAGA 764
Qy	601	GGATAC	CAGAGCAAGAGAGGAAGAGCAGAGCGCTTTGGGATTCGCTGATGAAAGT 660
Db	765	GGATAC	CAGAGCAAGAGAGGAAGAGCAGAGCGCTTTGGGATTCGCTGATGAAAGT 824
Qy	661	TCCTGA	TACTTTCTGTTCTCCAGTGTTTTCCATTTCTCTCTTCTTTGTTGTCACATATA 720
Db	825	TCCTGA	TACTTTCTGTTCTCCAGTGTTTTCCATTTCTCTCTTCTTTGTTGTCACATATA 884
Qy	721	TGCCTAA	TGACAGTCAATGTGCCTACGCTCTGCTCGCAATGAGGGAGCATGTACCCCA 780
Db	885	TGCCTAA	TGACAGTCAATGTGCCTACGCTCTGCTCGCAATGAGGGAGCATGTACCCCA 944
Qy	781	GGTACAT	CCATGAATCTGGCGAGCATTTTGACTTATTGCTGTTCAGCTTTTAAGGTGTT 840
Db	945	GGTACAT	CCATGAATCTGGCGAGCATTTTGACTTATTGCTGTTCAGCTTTTAAGGTGTT 1004
Qy	841	GTGTTTT	TGTTTTGATTAATGTGCTGCTTAAT 873
Db	1005	GTGTTTT	TGTTTTGATTAATGTGCTGCTTAAT 1037

[illegible]

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
	1									
	AUTHORS Yue, H., Tang, Y. T., Lai, P., Reddy, R., Batra, S., Baughn, M. R., Yang, J., Azimzai, Y., Lu, D. A., Au-Young, J., and Shih, L. L.									
	TITLE Full-length molecules expressed in human tissues									
JOURNAL	Patent: WO 0070047-A 86 23-NOV-2000;									
	Incyte Genomics, Inc. (US)									
FEATURES	Location/Qualifiers									
	1..1071									
ORIGIN	/organism="Homo sapiens"									
	/mol_type="unassigned DNA"									
	/db_xref="taxon:9606"									
	/note="Incyte ID No: 4093555CB1"									
Query Match 100.0%; Score 873; DB 6; Length 1071;										
Best Local Similarity 100.0%; Pred. No. 8.2e-205;										
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;										
Qy	1	TTGAGTTCAGGGGTAA	CAAGAATGCGACCGACGCTGGAGCTCCATAAGCTAAAGCTTATGC	60						
Db	165	TGAGTTCAGGGGTAA	CAAGAATGCGACCGACGCTGGAGCTCCATAAGCTAAAGCTTATGC	224						
Qy	61	CGAACTAAAGCAAGAAT	GTCTTCTCGTGGTTTGGAGACCAAGGGAATAAGCAAGATCT	120						
Db	225	CGAACTAAAGCAAGAAT	GTCTTCTCGTGGTTTGGAGACCAAGGGAATAAGCAAGATCT	284						
Qy	121	TATCCACAGACTCGAGGCATAT	CTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAAGA	180						
Db	285	TATCCACAGACTCCAGGCATAT	CTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAAGA	344						
Qy	181	TGTACTGGGAGATGAAC	CAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA	240						
Db	345	TGTACTGGGAGATGAAC	CAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA	404						
Qy	241	GGAAAGAACCCCTG	AAAAAACTGTGATGTGGCAGCAGAGAAGAAAGTGTGTAAAAATTAC	300						
Db	405	GGAAAGAACCCCTG	AAAAAACTGTGATGTGGCAGCAGAGAAGAAAGTGTGTAAAAATTAC	464						
Qy	301	ATCTGAATAACCA	CAGACTGAGAGAATGCAAGAGAGGGCTGAACGATTCATGTACTCTGT	360						
Db	465	ATCTGAATAACCA	CAGACTGAGAGAATGCAAGAGAGGGCTGAACGATTCATGTACTCTGT	524						
Qy	361	GAGCTTGGAGGTAA	GAAAGCTGCTCGGCAGCTAGTTTGGGATTTCTTCAGTTCACAAC	420						
Db	525	GAGCTTGGAGGTAA	GAAAGCTGCTCGGCAGCTAGTTTGGGATTTCTTCAGTTCACAAC	584						
Qy	421	AAAAGTCTCTCAT	CTGATAACAAACCTATGTGTTAACTTGGATAAGCTGAAGAAAAGAGC	480						
Db	585	AAAAGTCTCTCAT	CTGATAACAAACCTATGTGTTAACTTGGATAAGCTGAAGAAAAGAGC	644						
Qy	481	TCAAAGATTGGTTT	GAAATGCTCTTCAATCTCCAGAAAGCTTGAAGATGATGAGAAACT	540						
Db	645	TCAAAGATTGGTTT	GAAATGCTCTTCAATCTCCAGAAAGCTTGAAGATGATGAGAAACT	704						
Qy	541	GAATAAGAGAAAGG	CGGATTTGGGATTTGCAAGTTCAGCTGGAACTGGACACCAAGA	600						
Db	705	GAATAAGAGAAAGG	CGGATTTGGGATTTGCAAGTTCAGCTGGAACTGGACACCAAGA	764						
Qy	601	GGATACAGAGCG	CAAGAAAGAGACAGACAGCGCTTTGGGATTTGCCTGATGAAAAAGT	660						
Db	765	GGATACAGAGCG	CAAGAAAGAGAGACAGACAGCGCTTTGGGATTTGCCTGATGAAAAAGT	824						
Qy	661	TCCTGATACATTT	CTTGTTCTCCAGTGTTTCTCAATTTCTCTCTTCTTCTTGGTCACATATA	720						
Db	825	TCCTGATACATTT	CTTGTTCTCCAGTGTTTCTCAATTTCTCTCTTCTTCTTGGTCACATATA	884						
Qy	721	TGCCTAAATGCA	CAGTCATGTGCTACGCTCGCTCGCAATGAGGGAGCATGTACCCCA	780						
Db	885	TGCCTAAATGCA	CAGTCATGTGCTACGCTCGCTCGCAATGAGGGAGCATGTACCCCA	944						
Qy	781	GSTACATCCAT	GAACTGCGGCGACAGTTTGACATTATTGCTGTTTTCAGCTTTTAAGGTTGTT	840						

|||||
945 GGATACATCCATGAACTCGGAGAGATTGACATTAATGCTGTTTACGCTTTAAAGTTGTT 1004
Db
|||||
841 GTGTTTTGTTTTGATTATGTTGCTTGTAAAT 873
Qy
|||||
1005 GTGTTTTGTTTTGATTATGTTGCTTGTAAAT 1037
Db

RESULT 3
AX405619
LOCUS
DEFINITION
Sequence 34 from Patent WO0222660.
ACCESSION
AX405619
VERSION
AX405619.1 GI:21438677
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,
Xue, A.J., Yang, Y., Wehrman, T. and Drmanac, R.T.
TITLE
Novel nucleic acids and polypeptides
JOURNAL
Patent: WO 022660-A 34 21-MAR-2002;
HYSEQ, INC. (US)
FEATURES
Location/Qualifiers
source
1. 1520
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
636..1268
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD34726.1"
/db_xref="GI:21438678"
/translation="MATEVELHKLKLAELKQELARGLTKGIKODLIHRLQALYLE
HASEEANEEDVLDGETTEETKEFLPVRKEEPPEKTVDVAAEKVKVITSEIPQTER
MQRARFNVFVLESKKAARAFGISVPTKGLSSDNKPMVNLKDKRAORFGLN
VSSISRKSEDEDEKKRERFVITSSAGTGTTEDEAKKRKAERFGIA"
ORIGIN
Query Match 100.0%; Score 873; DB 6; Length 1520;
Best Local Similarity 100.0%; Pred. No. 8.3e-205;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGAGTGCAGGGGTAACAAGATGCGACCGAGCGTGGAGCTCCATAGCTTAAAGCTTGC 60
Db 617 TGGAGTGCAGGGGTAACAAGATGCGACCGAGCGTGGAGCTCCATAGCTTAAAGCTTGC 676
Qy 61 CGAACTAAAGCAAGAATGTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCT 120
Db 677 CGAACTAAAGCAAGAATGTTCTGCTCGTGGTTGGAGACCAAGGGAATAAGCAAGATCT 736
Qy 121 TATCCACAGACTCAGGCATATCTTGAAGAATGCTGAAGAGGAGGCAATGAAGAAGA 180
Db 737 TATCCACAGACTCAGGCATATCTTGAAGAATGCTGAAGAGGAGGCAATGAAGAAGA 796
Qy 181 TGTAAGTGGAGATGAACACAGAGGAAGAAGAAACCAAGCCATTCAGCTCCCTGTCAAGA 240
Db 797 TGTAAGTGGAGATGAACACAGAGGAAGAAGAAACCAAGCCATTCAGCTCCCTGTCAAGA 856
Qy 241 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAGAAAGTGGTGAATAATAC 300
Db 857 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAGAAAGTGGTGAATAATAC 916
Qy 301 ATCTGAAATACACAGACTGAGAGAATGCAAGAGGGCTGAACATTCATGATGACCTGT 360
Db 917 ATCTGAAATACACAGACTGAGAGAATGCAAGAGGGCTGAACATTCATGATGACCTGT 976
Qy 361 GAGCTTGGAGATGAAGAAGCTGCTCGGAGCTAGCTTGGGATTTCTTCAGTTCCAAC 420
Db 977 GAGCTTGGAGATGAAGAAGCTGCTCGGAGCTAGCTTGGGATTTCTTCAGTTCCAAC 1036

Qy 421 AAAAGCTCTGTCATCTGATTAACAAACCTATGTTTAACTTGGATAAGCTGAAGGAAGAGC 480
Db 1037 AAAAGCTCTGTCATCTGATTAACAAACCTATGTTTAACTTGGATAAGCTGAAGGAAGAGC 1096
Qy 481 TCAAAGATTTGGTTTGAATGTCCTTCAATCTCAGAAAGTCTGAAGATGATGAGAAACT 540
Db 1097 TCAAAGATTTGGTTTGAATGTCCTTCAATCTCAGAAAGTCTGAAGATGATGAGAAACT 1156
Qy 541 GAAAAAGAGGAAGAGGATTTGGGATTTGTCACAAAGTTACAGTTCAGCTGGAACTGGAACACAGA 600
Db 1157 GAAAAAGAGGAAGAGGATTTGGGATTTGTCACAAAGTTACAGTTCAGCTGGAACTGGAACACAGA 1216
Qy 601 GGATACAGAGGCAAGAAGAGGAAAGAGAGCAGAGCGCTTTGGGATTCCTGATGAAAGT 660
Db 1217 GGATACAGAGGCAAGAAGAGGAAAGAGAGCAGAGCGCTTTGGGATTCCTGATGAAAGT 1276
Qy 661 TCCTGATACATTTCTGTTTCTCCAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db 1277 TCCTGATACATTTCTGTTTCTCCAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1336
Qy 721 TGCTTAATGCAAGTCATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 1337 TGCTTAATGCAAGTCATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1396
Qy 781 GGTACATCCATGAAGTGGGAGCAGCTTTGACATTTATGCTGTTTACGCTTTAAGGTTGTT 840
Db 1397 GGTACATCCATGAAGTGGGAGCAGCTTTGACATTTATGCTGTTTACGCTTTAAGGTTGTT 1456
Qy 841 GTGTTTTGTTTTGATTATGTTGCTTGTAAAT 873
Db 1457 GTGTTTTGTTTTGATTATGTTGCTTGTAAAT 1489

RESULT 4
HSA409089
LOCUS
DEFINITION
Homo sapiens HCC-1 gene.
ACCESSION
AJ409089
VERSION
AJ409089.1 GI:13940309
KEYWORDS
HCC-1 gene; hepatocellular carcinoma; nuclear protein.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Choong, M.L., Tan, L.K., Lo, S.L., Ren, E.C., Ou, K., Ong, S.E.,
Liang, R.C., Sow, T.K. and Chung, M.C.
TITLE
An integrated approach in the discovery and characterization of a
novel nuclear protein over-expressed in liver and pancreatic tumors
JOURNAL
FEBS Lett. 496 (2-3), 109-116 (2001)
MEDLINE
21255661
PUBMED
11356193
REFERENCE
2 (bases 1 to 1557)
AUTHORS
Choong, M.L.
Direct Submission
TITLE
Submitted (21-FEB-2001) Choong M.L., National University of
Singapore, Bioprocessing Technology Centre, MD 11 Level 5, 10
Medical Drive, 117597, SINGAPORE
FEATURES
Location/Qualifiers
source
1. 1557
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q22"
/cell_line="HCC-M"
/cell_type="hepatocellular carcinoma"
/tissue_type="liver"
1. 1557
/genes="HCC-1"
1. 687
/genes="HCC-1"
/evidence="experimental"
gene
promoter

```
misc_feature 305..594
/gene="HCC-1"
/notes="non-functional upstream open reading frame"
/function="tightens and regulates translation of HCC-1
gene"
repeat_region 446..453
/rpt_type=DIRECT
mRNA 661..1557
/gene="HCC-1"
CDS 688..1320
/gene="HCC-1"
/protein="HCC-1 protein"
/codon_start=1
/product="HCC-1 protein"
/protein_id="CAC37950.1"
/db_xref="GI:13940310"
/db_xref="GOA:P82979"
/db_xref="UniProt/Swiss-Prot:P82979"
/translations="MATEVELKHLKLAELKQRLGLETKGIKODLHRLQVLEE
HAREEANEEDVLGDEETERTKPIELPVKEEPEKTVDAAEKKVKVKITSEIPQTER
MQKAEAEFNVPSLESKKAARAFGISVPTTKGLSDNKPWNLDKLKRAQRFGLN
VSSIRKSEDEKLRKRERFGLVTSAGTGTITTEAKRKRKAERFGIA"
3'UTR 1321..1557
/gene="HCC-1"
/evidence=experimental
repeat_region 1358..1365
/rpt_type=DIRECT
polya_site 1542
/gene="HCC-1"
/evidence=experimental
ORIGIN
Query Match 100.0%; Score 873; DB 9; Length 1557;
Best local Similarity 100.0%; Pred. No. 8.3e-205;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGAGTGAGGGGTAAACAAGATGGCGACGAGCGTGGAGCTCCATAAGCTAAAGCTTGC 60
DB 669 TGGAGTGAGGGGTAAACAAGATGGCGACGAGCGTGGAGCTCCATAAGCTTAAAGCTTGC 728
QY 61 CGAAGTAAAGCAAGAATGTCTTGCTGCTGTTGGAGCAAGGGGATAAAGCAAGATCT 120
DB 729 CGAAGTAAAGCAAGAATGTCTTGCTGCTGTTGGAGCAAGGGGATAAAGCAAGATCT 788
QY 121 TATCCACAGACTCCAGGCATATCTTGAAGACATGCTGAAGAGGAGGCAATGAAGAGA 180
DB 789 TATCCACAGACTCCAGGCATATCTTGAAGACATGCTGAAGAGGAGGCAATGAAGAGA 848
QY 181 TGTAAGTGGAGATGAACAGAGGAAGAAAGAAAGAAAGCCCATTTGAGCTCCCTGTCAAGA 240
DB 849 TGTAAGTGGAGATGAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 908
QY 241 GGAAGAACCCCTGAAAAAATCTGTGTGATGTGGCAGCAGAGAGAAAGTGTGAAATATAC 300
DB 909 GGAAGAACCCCTGAAAAAATCTGTGTGATGTGGCAGCAGAGAGAAAGTGTGAAATATAC 968
QY 301 ATCTGAATACCAAGACTGAGAGAAATGCGAGAGAGGAGGCTGAACGATTCATCTGCTGT 360
DB 969 ATCTGAATACCAAGACTGAGAGAAATGCGAGAGAGGAGGCTGAACGATTCATCTGCTGT 1028
QY 361 GAGCTTGGAGATGAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAAC 420
DB 1029 GAGCTTGGAGATGAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAAC 1088
QY 421 AAAAGGTCTGATCTGATTAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAGAGAGC 480
DB 1089 AAAAGGTCTGATCTGATTAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAGAGAGC 1148
QY 481 TCAAGATTTGGTTGATGCTCTTCAATCTCCAGAAAGCTTCAAGATGATGAGAACT 540
DB 1149 TCAAGATTTGGTTGATGCTCTTCAATCTCCAGAAAGCTTCAAGATGATGAGAACT 1208
QY 541 GAAAAAGAGGAAGAGGCGATTTGGGATTTGTACAAAGTTGACAGCTGGAACCTGGAACCAAGAG 600
|||||
```

```
DB 1209 GAAAAAGAGGAAGAGGCGATTTGGGATTTGTACAAAGTTGACGTGGAATCGGAACCAAGAG 1268
QY 601 GGATACAGAGGCAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
DB 1269 GGATACAGAGGCAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1328
QY 661 TCCTGATACCTTCTGTTCTCCAGTGTTTTTCCAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
DB 1329 TCCTGATACCTTCTGTTCTCCAGTGTTTTTCCAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1388
QY 721 TGCCTAAATGCACAGTCATGTGCTACGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 1389 TGCCTAAATGCACAGTCATGTGCTACGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1448
QY 781 GGTACATCCATGAACCTGCGGACGAGTTTGACTTATGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 1449 GGTACATCCATGAACCTGCGGACGAGTTTGACTTATGCTGTTGCTGTTGCTGTTGCTGTTGCTGTT 1508
QY 841 GTGTTTTTCTGTTTGAATATGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTT 873
DB 1509 GTGTTTTTCTGTTTGAATATGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTTGCTGTT 1541

RESULT 5
BC007099
LOCUS BC007099
DEFINITION Homo sapiens cytokine induced protein 29 kDa, mRNA (cDNA clone
MGC:14726 IMAGE:4273903), complete cds.
ACCESSION BC007099
VERSION BC007099.1 GI:13937970
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 923)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Frange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,
Schmerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 923)
Strausberg, R.
Direct Submission
Submitted (30-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
```

Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: iRAI, Plate: 21 Row: 9 Column: 18
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21361588.

FEATURES

Location/Qualifiers
1..923
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:14726 IMAGE:4273903"
/tissue type="Prostate"
/clone_lib="NIH MGC_83"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
1..923
/gene="CIP29"
/note="synonyms: MGC14726, HCC-1, HSPC316"
/db_xref="LocusID:84324"
22..654
/gene="CIP29"
/codon_start=1
/product="cytokine induced protein 29 kDa"
/protein_id="AAH07099.1"
/db_xref="GI:13937971"
/db_xref="LocusID:84324"
/translation="MATETVELHKLAELKQECIARGLFTKGIKQDLIHLRLQYLBE
HAEEANEEDVLGDETEETKPIELPVKEEPEKTVDAAEKKVVKITSEIPQTER
MOKAERFNVSLESKAAARAFGISVPTKGLSDNKPWNLDLKKRAORFLN
VSSISRSKEDDEKLKKRERFGIVTSSAGTGTEDTEAKKKRAERFGIA"

gene

CDS

ORIGIN

Query Match 99.8%; Score 871.4; DB 9; Length 923;
Best Local Similarity 99.9%; Pred. NO. 2e-204;
Matches 872; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGAGTGGGGTACAGATGGGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGC 60
Db 3 TGGAGTGGGGTACAGATGGGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGC 62

Qy 61 CGAACTAAAGCAAGATCTTCTGCTCGTGGTTTGGAGACCAAGGAATAAAGCAAGATCT 120
Db 63 CGAACTAAAGCAAGATCTTCTGCTCGTGGTTTGGAGACCAAGGAATAAAGCAAGATCT 122

Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAAGA 180
Db 123 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAAGA 182

Qy 181 TGTACTGGAGATGAACAGAGGAAGAAAGAAACAAAGCCCATTTAGCTCCCTGTCAAGAAGA 240
Db 183 TGTACTGGAGATGAACAGAGGAAGAAAGAAACAAAGCCCATTTAGCTCCCTGTCAAGAAGA 242

Qy 241 GGAAGACCCCTGAAAAAATGTTGTGTCGACGACAGAGGAAGAGTGTGAAATTTAC 300
Db 243 GGAAGACCCCTGAAAAAATGTTGTGTCGACGACAGAGGAAGAGTGTGAAATTTAC 302

Qy 301 ATCTGAAATACACAGACTGAGAGAAATGCAGAAAGAGGCTGAAACGATTCAATGTACCTGT 360
Db 303 ATCTGAAATACACAGACTGAGAGAAATGCAGAAAGAGGCTGAAACGATTCAATGTACCTGT 362

Qy 361 GAGCTTGAGAGTGAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAC 420
Db 363 GAGCTTGAGAGTGAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAC 422

Qy 421 AAAAGGCTGTCTGATGATACAAACCTATGGTTAACTTGGATAGCTGAGAGGAAGAGC 480
Db 423 AAAAGGCTGTCTGATGATACAAACCTATGGTTAACTTGGATAGCTGAGAGGAAGAGC 482

Qy 481 TCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGCTCTGAAGATGATGAGAACT 540
Db 483 TCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGCTCTGAAGATGATGAGAACT 542

Qy 541 GAAAGAGGAGGAGGAGGATTTGGGATTTGTCAAGTTTCAAGTGGAACTGGAACACACAGA 600
Db 543 GAAAGAGGAGGAGGAGGATTTGGGATTTGTCAAGTTTCAAGTGGAACTGGAACACACAGA 602

Qy 601 GGATACAGAGCAAGAAAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCGTATGAAAGT 660
Db 603 GGATACAGAGCAAGAAAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCGTATGAAAGT 662

Qy 661 TCCTGATACCTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCATATA 720
Db 663 TCCTGATACCTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCATATA 722

Qy 721 TGCCTAAATGCACAGTCATGTCCTACGCTGCTGCTCGCAATGAGGAGCATGTACCCCA 780
Db 723 TGCCTAAATGCACAGTCATGTCCTACGCTGCTGCTCGCAATGAGGAGCATGTACCCCA 782

Qy 781 GGATACATCCATGAACCTGCGCAGCAGTTTGAATTTGCTGTTTCAAGTTTAAAGTGT 840
Db 783 GGATACATCCATGAACCTGCGCAGCAGTTTGAATTTGCTGTTTCAAGTTTAAAGTGT 842

Qy 841 GTGTTTTGTTTTGATTTGCTTGTGTTAAT 873
Db 843 GTGTTTTGTTTTGATTTGCTTGTGTTAAT 875

RESULT 6
CQ413602 LOCUS CQ413602 2553 bp DNA linear PAT 23-JAN-2004
DEFINITION Sequence 20673 from Patent WO0170979.
ACCESSION CQ413602
VERSION CQ413602.1 GI:41321383
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
AUTHORS Lee, J. and Lillie, J.
TITLE Genes, compositions, kits, and method for identification,
assessment, prevention, and therapy of ovarian cancer
JOURNAL Patent: WO 0170979-A 20673 27-SEP-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source Location/Qualifiers
1..2553
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 79.1%; Score 690.8; DB 6; Length 2553;
Best Local Similarity 99.7%; Pred. No. 8.5e-160;
Matches 692; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 93 TGGAGACCAAGGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAC 152
Db 715 TGGAGACCAAGGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAC 774

Qy 153 ATGCTGAAGAGGAGGCAAAATGAAGAGATGTACTTGGGAGATGAACAGAGGAAGAGAAA 212
Db 775 ATGCTGAAGAGGAGGCAAAATGAAGAGATGTACTTGGGAGATGAACAGAGGAAGAGAAA 834

Qy 213 CAAAGCCCATTTAGCTCCCTGTCMAAGAGGAAGAACCCCTCTGAAGAACTGTTGATGTGG 272
Db 835 CAAAGCCCATTTAGCTCCCTGTCMAAGAGGAAGAACCCCTCTGAAGAACTGTTGATGTGG 894

Qy 273 CAGCAGAGGAAGAGTGGTGAATTAATCATCTGAAATATACATCTGAAATATACACAGACTGAGAGAAATGCGA 332
Db 895 CAGCAGAGGAAGAGTGGTGAATTAATCATCTGAAATATACATCTGAAATATACACAGACTGAGAGAAATGCGA 954

```

QY 333 AGAGGGCTGAACGATTCAATGTACCTGTGAGCTTGAGAGTAAGAAGCTGCTCGGGCAG 392
Db |||||
QY 955 AGAGGCTGACCATTCATGTACCTGTGAGCTTGAGAGTAAGAAGCTGCTCGGGCAG 1014
Db |||||
QY 393 CTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAGGTCTGTCATCTGATACAAACCTATGG 452
Db |||||
QY 1015 CTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAGGTCTGTCATCTGATACAAACCTATGG 1074
Db |||||
QY 453 TTAACCTGGATAGCTGAAGGAAGAGCTCAAGAGTTGGTTGAAATGCTCTTCAATCT 512
Db |||||
QY 1075 TTAACCTGGATAGCTGAAGGAAGAGCTCAAGAGTTGGTTGAAATGCTCTTCAATCT 1134
Db |||||
QY 513 CCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAGCGCATTTGGGATTTGTC 572
Db |||||
QY 1135 CCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGAGGAGCGCATTTGGGATTTGTC 1194
Db |||||
QY 573 CAAGTTCACTGGAATCGGAACTGGACACAGAGGATACAGAGGCAAGAGAGGAAAAAGAGCAG 632
Db |||||
QY 1195 CAAGTTCACTGGAATCGGAACTGGAAACACAGAGGATACAGAGGCAAGAGAGGAAAAAGAGCAG 1254
Db |||||
QY 633 AGCGTTTGGGATTCGCTGATGAAAGTTCCCTGATATCTTCTCTCCAGTGTCTTCC 692
Db |||||
QY 1255 AGCGTTTGGGATTCGCTGATGAAAGTTCCCTGATATCTTCTCTCCAGTGTCTTCC 1314
Db |||||
QY 693 TTCTCTCTCTCTCTCTCTCTCTGTCATATATGCTTAAATGACAGTGCATGTGCCTACGTCCT 752
Db |||||
QY 1315 TTCTCTCTCTCTCTCTCTCTCTGTCATATATGCTTAAATGACAGTGCATGTGCCTACGTCCT 1374
Db |||||
QY 753 GCCTCGCAATGAGGAGCATGTATCCCCAGGTACA 786
Db |||||
QY 1375 GCCTCGCAATGAGGAGCATGTATCCCCAGGTACA 1408
Db |||||

RESULT 7
AC073063
LOCUS
DEFINITION Homo sapiens BAC clone RP11-136B3 from 7, complete sequence.
AC073063
VERSION
KEYWORDS HTG.
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE
AUTHORS Armstrong,J., Maupin,R. and Long,N.
TITLE The sequence of Homo sapiens BAC clone RP11-136B3
JOURNAL Unpublished (2001)
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

```

TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 84113)
Waterston,R.
Direct Submission
Submitted (04-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 18, 2001 this sequence version replaced gi:13431102.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiense@wustl.wustl.edu
----- Summary Statistics

Center project name: H_NH0136B03

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/FTP/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-114P12, 2000 bp overlap;
the clone sequenced to the left is GSI-259H13, 200 bp overlap.
Actual start of this clone is at base position 96082 of GSI-259H13;
actual end is at base position 84113 of RP11-136B3.

FEATURES

source	Location/Qualifiers
1. .84113	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7"
	/clone="RP11-136B3"
	/clone_lib="RP11-11"
22. .56	/rpt_family="AT-rich"
46. .330	/rpt_family="Alu"
335. .405	/rpt_family="MER1_type"
545. .613	/rpt_family="MER1_type"

```
repeat_region 614..780 /rpt_family="Alu"
repeat_region 781..1078 /rpt_family="Alu"
repeat_region 1080..1114 /rpt_family="Alu"
repeat_region 1115..1406 /rpt_family="MER1_type"
repeat_region 1407..1540 /rpt_family="Alu"
misc_feature 1833..2211 /rpt_family="MER1_type"
misc_feature 2322..2394 /note="match to EST BE147565 (NID:g8610289)"
repeat_region 2337..2489 /rpt_family="Alu"
repeat_region 2497..2619 /rpt_family="WIR"
repeat_region 2620..2644 /rpt_family="WIR"
repeat_region 2926..2991 /rpt_family="(T)n"
repeat_region 3055..3355 /rpt_family="WIR"
repeat_region 3374..3395 /rpt_family="Alu"
repeat_region 3552..3852 /rpt_family="(TTTTTA)n"
repeat_region 3853..3972 /rpt_family="Alu"
repeat_region 4009..4388 /rpt_family="Alu"
repeat_region 4419..4731 /rpt_family="R2_type"
repeat_region 4786..5085 /rpt_family="Alu"
repeat_region 5203..5238 /rpt_family="L1"
repeat_region 5239..5651 /rpt_family="L1"
repeat_region 5642..5663 /rpt_family="(TAAA)n"
misc_feature 5885..6027 /note="match to EST BE795688 (NID:g10216886)"
misc_feature 5887..6025 /note="match to EST BG722929 (NID:g14002116)"
misc_feature 5887..6025 /note="similar to EST BG917776 (NID:g14298252)"
misc_feature 5990..6025 /note="match to EST BG748931 (NID:g14059584)"
repeat_region 6106..6382 /rpt_family="L2"
misc_feature 6330..6334 /note="match to EST BF894562 (NID:g12286021)"
repeat_region 6383..6682 /rpt_family="Alu"
repeat_region 6683..6744 /rpt_family="L2"
repeat_region 6745..7046 /rpt_family="Alu"
repeat_region 7017..7042 /rpt_family="(GAAA)n"
repeat_region 7047..7139 /rpt_family="L2"
misc_feature 7251..7849 /note="match to EST AV717144 (NID:g10814296)"
repeat_region 7435..7732 /rpt_family="Alu"
repeat_region 7735..8052 /rpt_family="Alu"
misc_feature 7907..8001 /note="match to EST AV717144 (NID:g10814296)"
misc_feature 7927..8032 /note="match to EST AV717144 (NID:g10814296)"

repeat_region 8017..8086 /note="match to EST BE147565 (NID:g8610289)"
repeat_region 8118..8234 /rpt_family="A-rich"
repeat_region 8219..8241 /rpt_family="Alu"
repeat_region 8245..8543 /rpt_family="AT-rich"
misc_feature 8300..8302 /note="match to EST AV717144 (NID:g10814296)"
repeat_region 8713..9008 /rpt_family="Alu"
repeat_region 9055..9190 /rpt_family="Alu"
repeat_region 9191..9501 /rpt_family="Alu"

Query Match 74.2%; Score 647.4; DB 9; Length 84113;
Best Local Similarity 90.4%; Pred. No. 4.9e-149;
Matches 782; Conservative 0; Mismatches 46; Indels 37; Gaps 7;

Qy 7 GAGGGTAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGCCGAAC 66
Db 74969 GAGGGTAACAAGATGGTAACCTGAGATGGTGGAGCTCCATAAGCTGAAGCTTGCTAA 75028
Qy 67 AAAGCAAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 126
Db 75029 AAAGCAAGATGCTCTT-CTTGTGCTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCA 75087
Qy 127 CAGACTCCAGGCATATCTTGAAGCAATGCTGAAGAGGAGGCAAAATGAAGAAGATGCTACT 186
Db 75088 CAGACTCCAGGCATATCTTGAAGCAATGCTGAAGAGGAGGCAAAAT---GAAGATGCTACT 75144
Qy 187 GGGAGATGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246
Db 75145 AGGAGATGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75191
Qy 247 ACCCCCTG-AAAAAATGTTGATGTGGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 305
Db 75192 ACCCCCTGAAAAAATGTTGATGTGGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75251
Qy 306 AAATACACAGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365
Db 75252 AAATACACAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75311
Qy 366 TGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTTCCACAAAG 425
Db 75312 TGGAGAGTAAGAAAGCTGCTCAGCAGCTAGGTTTGGGATTTCTTCAGTTTCCACAAAG 75370
Qy 426 GTCTGTCATCTGATAACAAACCTATGTTAACTTGGATTAAGCTGAAGAGAGAGAGAGAGAGAG 485
Db 75371 GTCTGTCATCTG- AACACACCTATGTTAACTTGGATTAAGCTGAAGAGAGAGAGAGAGAG 75428
Qy 486 GATTGGTTTGAATGCTCTTCAATCTCCAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545
Db 75429 GATTGGTTTGAATGCTCTTCAATTTCCAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 75488
Qy 546 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605
Db 75489 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75548
Qy 606 CAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 665
Db 75549 CAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75608
Qy 666 ATACTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 725
Db 75609 ATACTTCTGTTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 75668
Qy 726 AAATGCACAGTCATGTGCGCTACGTTCCTGCTCGCAATGAGGGAGAGAGAGAGAGAGAGAGAGAG 785
Db 75669 AA-----TGCAGTCTTGCCTTGCATGAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75712
```

QY 786 ATCCATGAACGGGACGAGTTTGACTTATGCTTTTTCAGCTTTAAAGGTGTGTGTT 845
 |||||
 Db 75713 AACTGTGAACGGCGGCGAGTTTGACTTATGCTGTTTCAGCTTTAAAGGTGTGTGTT 75772
 |||||
 QY 846 TTTGTTTTGATTATGTTGCTTGT 870
 |||||
 Db 75773 TTTGTTTTGATTATGTTGCTTGT 75797
 |||||

RESULT 8
 AC146422/c
 LOCUS Pan troglodytes chromosome 7 clone RP43-27K9, complete sequence.
 DEFINITION AC146422 189507 bp DNA linear PRI 06-OCT-2004
 ACCESSION AC146422
 VERSION AC146422.4 GI:53828900
 KEYWORDS HTG.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 REFERENCE 1. (bases 1 to 189507)
 AUTHORS Wilson,R.K.
 TITLE The sequence of Pan troglodytes clone
 JOURNAL Unpublished
 REFERENCE 2. (bases 1 to 189507)
 AUTHORS Wilson,R.K.
 TITLE Direct Submission
 JOURNAL Submitted (13-AUG-2003) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 3. (bases 1 to 189507)
 AUTHORS Wilson,R.K.
 TITLE Direct Submission
 JOURNAL Submitted (27-APR-2004) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 4. (bases 1 to 189507)
 AUTHORS Wilson,R.K.
 TITLE Direct Submission
 JOURNAL Submitted (06-OCT-2004) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA
 COMMENT On Oct 6, 2004 this sequence version replaced gi:46576146.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu
 Contact: submissions@wustl.edu
 ----- Project Information -----
 Center project name: C PT027K09
 ----- Mapping Information -----
 The position of this clone was established as part of a
 collaboration between the Chimpanzee Chromosome Y Mapping Project
 (Jennifer F. Hughes, Tatiana Pyntikova, Helen Skaletsky, Steve
 Rozen, and David C. Page at the Whitehead Institute for Biomedical
 Research, Cambridge MA) and the Washington University Genome
 Sequencing Center, St. Louis MO.

FEATURES
 source Location/Qualifiers
 1. 189507
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /chromosome="7"
 /clone="RP43-27K9"

ORIGIN
 Query Match 72.6%; Score 633.6; DB 9; Length 189507;
 Best Local Similarity 89.3%; Pred. No. 1.3e-145;
 Matches 775; Conservative 0; Mismatches 54; Indels 39; Gaps 7;
 QY 7 GAGGGGTAAACAAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTAAAGCTTCCGAACT 66
 |||||
 Db 132824 GAGGGGTAAACAAGATGGCTAATGAGATGGTGGAGCTCCGTAAGCTGTAAGCTTGTCTAAACT 132765
 |||||

QY 67 AAAGCAAGAATGCTTTGCTGCTGGTTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCA 126
 |||||
 Db 132764 AAAGCAAGAATGCTTTGCTGCTGGTTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCA 132705
 |||||
 QY 127 CAGACTCCAGGCATATCTTTGAAGAACATGCTGAAGAGAGGCAAAATGAAGAAGATGACT 186
 |||||
 Db 132704 CAGACTCCAGGCATATCTTTGAAGAACATGCTGAAGAGAGGCAAAAT--GAAGATGACT 132648
 |||||
 QY 187 GGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGTCTCCCTGTCAAGAGGAAGA 246
 |||||
 Db 132647 AGGAGATGAACAGAGGAAGAAACAA-----AACCCCTTTGTCAAATAGGAAGA 132601
 |||||
 QY 247 ACCCCCTG-AAAAACTGTTGATGTGGCAGCAGAGAAAGGTGGAATTTACATCTG 305
 |||||
 Db 132600 ACCCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGGTGGAATTTACATCTG 132541
 |||||
 QY 306 AAATACCACAGACTGAGAGAATCGAAGAGAGGCTCAACGATTTCAATGTACCTGTGAGCT 365
 |||||
 Db 132540 AAATACCACAGGCTGAGAGAATCGAAGAGAGGCGCAACAATTCAGTGTACCTGTGAGCT 132481
 |||||
 QY 366 TGGAGAGTAAAG---AAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAACA 422
 |||||
 Db 132480 TGGAGAGTAAAGTCAAGTAAATTCAGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAACA 132421
 |||||
 QY 423 AAGGCTGTCATCTGATTAACAAACCTATGTTAACTTGAAGCTGAAGGAAAGAGCTC 482
 |||||
 Db 132420 AA-GTCTGTCATCTG---AACACACCTATGTTAACTTTGGATAGCCGAAAGAAAGAGCTC 132364
 |||||
 QY 483 AAAGATTTTGGTTGAATGTCCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGA 542
 |||||
 Db 132363 AAAGATTTTGGTTGAATGTCCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGA 132304
 |||||
 QY 543 AAAAGAGGAAGAGCGGATTTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAACACAGAGG 602
 |||||
 Db 132303 AAAAGAGGAAGAGCGGATTTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAACACAGAGG 132244
 |||||
 QY 603 ATACAGAGCAAGAGAGAGGAAAGAGAGCGCTTTGGGATTTGGGATTTGCTGATGAAAGTTTC 662
 |||||
 Db 132243 ATACAGAGCAAGAGAGAGGAAAGAGAGCGCTTTGGGATTTGGGATTTGCTGATGAAAGTTTC 132184
 |||||
 QY 663 CTGATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 722
 |||||
 Db 132183 TTGATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 132124
 |||||
 QY 723 CTTAAATGCACAGTCATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 782
 |||||
 Db 132123 CTTAAA-----TGCACTTTGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCT 132080
 |||||
 QY 783 TACATCCATGAACGCGGAGCAGTTTGACTTATTTGCTGCTTTTTCAGCTTTTAAAGGTTGTTGT 842
 |||||
 Db 132079 TAAACTGTGAACGCGGAGCAGTTTGACTTATTTGCTGCTTTTTCAGCTTTTAAAGGTTGTTGT 132020
 |||||
 QY 843 GTTTTGTGTTTTCATTTATGTTGCTTGT 870
 |||||
 Db 132019 GTTTTGTGTTTTCATTTATGTTGCTTGT 131992
 |||||

RESULT 9
 AF486281
 LOCUS Homo sapiens proliferation associated cytokine-inducible protein
 DEFINITION CIP29 mRNA, complete cds.
 ACCESSION AF486281
 VERSION AF486281.1 GI:20069115
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1. (bases 1 to 633)
 AUTHORS Fukuda,S., Wu,D.W., Stark,K. and Pelus,L.M.
 TITLE Cloning and characterization of a proliferation-associated
 cytokine-inducible protein, CIP29

JOURNAL Biochem. Biophys. Res. Commun. 292 (3), 593-600 (2002)
MEDLINE 21920340
PUBMED 11922608
REFERENCE 2 (bases 1 to 633)
AUTHORS Fukuda,S., Wu,D.W., Stark,K. and Pelus,L.M.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2002) Microbiology/Immunology, Indiana University, School of Medicine, 1044 West Walnut Street, Indianapolis, IN 46202-5121, USA
FEATURES
source location/Qualifiers
1. .633
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="primary human breast cancer"
1. .633
/note="nuclear protein with SAP motif"
/codon_start=1
/product="proliferation associated cytokine-inducible protein CIP29"
/protein_id="AA09686.1"
/db_xref="GI:20069116"
/translation="MATETVELHLKLAEKQELARGLETGKIKODLIHRLQAYLEE
HAEAEEDVLGDEETETKPIELPKVEEPKTVDVAAEKVKVITSEIPOTER
MQRAERNPVVLESKARAARFGISSVPTKGLSDNPKPMWNLDKLKERARFGLN
VSSISRSKSEDEKLKKRKRPGIVTSSAGTGTTEDTEAKKKRAERFGIA"
ORIGIN
Query Match 72.5%; Score 633; DB 9; Length 633;
Best Local Similarity 100.0%; Pred. No. 1.5e-145;
Matches 633; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 20 ATGGCGACCGAGACGGTGGAGCTCATAAGCTAAAGCTTGCAGAACTAAAGCAAGAATGT 79
Db 1 ATGGCGACCGAGACGGTGGAGCTCATAAGCTAAAGCTTGCAGAACTAAAGCAAGAATGT 60
Qy 80 CTTGCTCGTGTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCA 139
Db 61 CTTGCTCGTGTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCA 120
Qy 140 TATCTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAGATGCTGGGAGATGAACA 199
Db 121 TATCTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAGATGCTGGGAGATGAACA 180
Qy 200 GAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTCTCAAGAGGAGAAACCCCTCGAAAAA 259
Db 181 GAGGAAGAGAAACAAAGCCCATTTGAGCTCCCTCTCAAGAGGAGAAACCCCTCGAAAAA 240
Qy 260 ACTGTTGATGGCAGCAGAGAGAAAGTGGTGAATAATTAATCTGAAATACCAAGACT 319
Db 241 ACTGTTGATGGCAGCAGAGAGAAAGTGGTGAATAATTAATCTGAAATACCAAGACT 300
Qy 320 GAGGAATGCAGAGAGGCTGAACGATTCATCTGCTGAGCTTGGAGGTAAAGAA 379
Db 301 GAGGAATGCAGAGAGGCTGAACGATTCATCTGCTGAGCTTGGAGGTAAAGAA 360
Qy 380 GCTGCTCGGCAGCTAGTGTGGGATTTCTTCAGTTCACAAAGAGTCTCTCATCTGAT 439
Db 361 GCTGCTCGGCAGCTAGTGTGGGATTTCTTCAGTTCACAAAGAGTCTCTCATCTGAT 420
Qy 440 AACAAACCTATGTTAACTTCGATAAGCTGAAGAAAGAGCTCAAGATTTGGTTTGAAT 499
Db 421 AACAAACCTATGTTAACTTCGATAAGCTGAAGAAAGAGCTCAAGATTTGGTTTGAAT 480
Qy 500 GTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAGAGAGAGGCGGA 559
Db 481 GTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAGAGAGAGGCGGA 540
Qy 560 TTTGGGATTGTCAACAGTTTCACTGGAACCTGGAACCCACAGAGGATACAGAGGCAAGAG 619
Db 541 TTTGGGATTGTCAACAGTTTCACTGGAACCTGGAACCCACAGAGGATACAGAGGCAAGAG 600
Qy 620 AGGAAAAGACAGACGCGCTTTGGGATTGCTGTA 652

Db 601 AGGAAAAGACAGACGCGCTTTGGGATTGCTGTA 633
AC145981 220280 bp DNA linear HTG 01-AUG-2003
Pan troglodytes chromosome UNK clone RP43-13G23, *** SEQUENCING IN
PROGRESS ***, 46 unordered pieces.
AC145981
AC145981.1 GI:33386932
HTG; HTGS PHASE1.
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
Wilson,R.K.
The sequence of Pan troglodytes clone
Unpublished
2 (bases 1 to 220280)
Wilson,R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: C.PT013G23

* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1338: contig of 1338 bp in length
* 1339: gap of unknown length
* 1439: contig of 1199 bp in length
* 2637: gap of unknown length
* 2737: gap of unknown length
* 4065: contig of 1328 bp in length
* 4165: gap of unknown length
* 5254: contig of 1089 bp in length
* 5354: gap of unknown length
* 5355: contig of 1724 bp in length
* 7079: gap of unknown length
* 7179: contig of 1093 bp in length
* 8272: gap of unknown length
* 8371: contig of 1145 bp in length
* 9516: gap of unknown length
* 9617: contig of 1329 bp in length
* 10946: gap of unknown length
* 11046: contig of 1819 bp in length
* 12865: gap of unknown length
* 12965: contig of 2053 bp in length
* 15018: gap of unknown length
* 15117: contig of 1576 bp in length
* 16694: gap of unknown length
* 16794: contig of 1921 bp in length
* 18715: gap of unknown length
* 18815: contig of 2518 bp in length
* 21332: gap of unknown length
* 21433: contig of 1808 bp in length
* 23241: gap of unknown length
* 23341: contig of 1592 bp in length
* 25032: gap of unknown length
* 27092: contig of 2060 bp in length


```

Qy 481 TCAAGATTGGTTGAATGCTCTTCAATCTCAGAAAGTCTGAAGATGATGAGAACT 540
Db 27556 ACAGAGATTGGTTGAATGCTCTTCCATCTCTAGAAAGTCTGAGGATGATGAGAAGCT 27615
Qy 541 GAAAAGAGGAGGAGCGATTGGGATTGTGCACAAAGTTTCAGCTGGAAGCTGGAACACACAGA 600
Db 27616 GAAGAACGAGAGAGATTGGGATTGTGCACAAAGTTCAGCTGGAAGCTGGAACACACAGA 27675
Qy 601 GGATACAGAGCAAGAGAGAGAAAGAGAGAGCGCTTTGGGATTGCTCATGAAAGT 660
Db 27676 GGATACAGAGCAAGAGAGAGAAAGAGAGAGCGCTTTGGGATTGCTCATGAAAGT 27735
Qy 661 TCCTGATACCTTCTGCTTCAGTGTCTTCATTCCTCTCTCTCTCTCTCTCTCTCTCATATATA 720
Db 27736 TCGTG---CTTCTGCACCATAGTGTCTTCCATTTCTCAAAATTTCTCTGTTGTTTATATA 27792
Qy 721 TGCCTAAATGACAGTCATGCGCTACGTC-----CTGCCCTCGCAATGAG 765
Db 27793 TACATACACACCCCTACCTAGTATCTATCTACATACACAGTCATGTGCTTCACAGTGAG 27852
Qy 766 GGAGCATGTATCCCGAGGTACATCCATGAAGTTCGCGCAGCAGTTTGACTTTATGCTGTTTC 825
Db 27853 GAAGCATGTACGTTGGTGTACATCCCTGATTCAGTACGAGTGTATTTAACTGCTCTTCT 27912
Qy 826 AGCTTTAAGTGTGCTGCTTTGTTGTTTGTATGTT 863
Db 27913 GCGTTTAAGATTGTTGCTGTTTGTGTTTGTATTTTAAATTT 27950

```

RESULT 12

```

AC132088/c
LOCUS AC132088 159681 bp DNA linear ROD 27-NOV-2003
DEFINITION Mus musculus BAC clone RP24-410M8 from chromosome 19, complete
sequence.
ACCESSION AC132088
VERSION AC132088.4 GI:34740419
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 159681)
AUTHORS Harkins,R. and Bielicki,L.
TITLE The sequence of Mus musculus BAC clone RP24-410M8
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 159681)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 159681)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 159681)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 159681)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 159681)
AUTHORS Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Sep 16, 2003 this sequence version replaced gi:33187079.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC

```

Web site: <http://genome.wustl.edu>
 Contact: submissions@wustl.edu
 ----- Summary Statistics
 Center project name: M_BB0410M08

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC124557 and AC134830.

FEATURES	Location/Qualifiers
source	1..159681 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /chromosome="19" /map="19" /clone="RP24-410M8" /clone_lib="RPCI-24" 1..405 /rpt_family="ERVVR" 504..738 /rpt_family="MIR" 1016..1148 /rpt_family="B4" 1264..1691 /rpt_family="MaLR" 1696..1839 /rpt_family="MER2_type" 1912..2118 /rpt_family="L1" 2160..2247 /rpt_family="MIR" 2643..2829 /rpt_family="B2" 2831..3011 /rpt_family="B2" complement(2934..3006) /product="tRNA-Ser" /note="Likely pseudogene (HMM SC=29.93 / Sec struct SC=-6.06)" 4669..4878 /rpt_family="B2" 5508..5655 /rpt_family="RMR1B" 5743..5839 /rpt_family="L1" 5924..6010 /rpt_family="RMR1B"
repeat_region	repeat_region
repeat_region	repeat_region
repeat_region	repeat_region
repeat_region	repeat_region
repeat_region	repeat_region
repeat_region	repeat_region
repeat_region	repeat_region
trna	trna
repeat_region	repeat_region
repeat_region	repeat_region
repeat_region	repeat_region
repeat_region	repeat_region

		/mol_type="mRNA" /db_xref="taxon:10090" /clone="MGC:41057 IMAGE:1382354" /tissue_type="mammary gland, lactating mouse" /clone_lib="Soares_mammary_gland_NLMG" /lab_host="DH10B" /note=vector: pT7T3-Pac"
gene	1..908	
CDS	5..637	
	/gene="l110005A23Rik" /codon_start=1 /product="RIKEN cDNA l110005A23" /protein_id="AAH27510.1" /db_xref="GI:20380578" /db_xref="LocusID:66118" /db_xref="WGI:1913368"	
	16	
	/translat:ion="MAAETVELHKLAELKQBCLARGLETKIKODLINRLQAYLED HAEEANFEVDLGDETEEPFKPIELFVKEEPEPKAVDMASEKKVKVTISGPQTER MQKRAENFPVPSLESKKKAARAFGISVPTKGLSDTKPMVLNLDKLKERAQRFGLN VSSISRKSDEDEKLKRKRFPGIVTSAGTGTTEDTEAKKRKRAERFGIA"	
ORIGIN		
	Query Match	67.1%; Score 586.2; DB 10; Length 908;
	Best Local Similarity	81.9%; Pred. No. 5.7e-134;
	Matches	707; Conservative 0; Mismatches 138; Indels 18; Gaps 2;
Qy	16	CAAGATCGCACCGACGGTGGAGCTCCATAAGCTTAAGCTTCGCCGAACCTAAAGCAAGA 75
Dd	1	CAAGATGGCGCCGAGACGGTGGAGCTCCACAAGCTGAAGCTTGCTGAACCTTAAGCAAGA 60
Qy	76	ATGTCTTGTCTGCTGGTGTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGACTCCA 135
Dd	61	ATGTCTTGTCTGCTGGTGTAGAGACCAAGGGAATAAACAAGATCTTATAATAGGCTGCA 120
Qy	136	GGCATATCTTGAAGAACATGCTGAAGAGGAGGCGAAATGAAGAGATGTACTGGAGATGA 195
Dd	121	GGCCTATCTTGAAGACCATTGAGAAAGAGAAAGCAAAATGAAGAGATGTACTGGAGATGA 180
Qy	196	AACAGAGGAAGAAGAACCAAGGCCATTGAGCTCCCTGCTCAAAGAGGAAGAACCCCCTGA 255
Dd	181	AACTGAGGAAGAAGAACCAAGCCCCATAGAACTGCCTGTTAAAGAGGAAGAACTCTCTGA 240
Qy	256	AAAAACTGTTGATGTGGCAGCAGAGAAGAAGTGGTGAATAATTACATCTGAATAACCACA 315
Dd	241	AAAGGCTGTTGATATGGCATCGGAAGAAGAGTGGTGAAGATTACATCTGGGATACCTCA 300
Qy	316	GACTTGAGAGAAATGCAGAAAGGGCTGAACGAAATCAATGTACTGTGAGCTTGGAGAGTAA 375
Dd	301	GACTTGAAGAAGATGCAGAAAGGGCTGAACGAAATCAATGTACTGTAAAGCCCTGGAGAGTAA 360
Qy	376	GAAGAGCTGCTCGGCGCAGCTAGTTTTGGATTTCTTTCAAGTTCACAAAAGGTCTGTCACTC 435
Dd	361	GAAGGCTGCCCGCGCGGCTAGTTTTGGAATTTCTTCAGTTCCAAACAAGGTTTATCATC 420
Qy	436	TGATAACAAACCTATGGTTAACTTTGGATAAGCTGAAGGAAGAGCTCAAAAGATTGGTTT 495
Dd	421	TGACACCAAGCCTATGGTTAACTTACCTGGATAAACTAAAGGAAGAGSCACAGAGATTGGTTT 480
Qy	496	GAATGTCTCTCAATCTCCAGAAAGCTGTAAGATGATGAGAAACTGAAAAAGAGGAAGGA 555
Dd	481	GAATGTCTCTTCCATCTCTAGAAAGTCTGAGGATGATGAGAAGCTGAAGAAACCGGAAGGA 540
Qy	556	GCGATTTTGGGATCTGTCAACAGTTCAGCTTGGAACTGGAAACCAACAGAGATACAGAGGCAAA 615
Dd	541	GAGATTTTGGGATTTGTGCAAGTTTCAGCTTGGAACTTGGAAACCAACAGAGATACAGAGGCAAA 600
Qy	616	GAAGAGAAAGAGCAGAGCGCTTTGGGATTTGCCCTGATGAAAAAGTTCTTGATCTTTCTG 675
Dd	601	GAAGAAGAAAGAGCAGAGCGCTTTTGGAAATTTGCATAATGAAAGTTCTGTG--CTTCTG 657

QY	676	TTCTCAGTGTGTTTTCAATTTCTCTCTCTCTTGTCACATATATGCCCTAAATGCACAG	735
DB	658	CACCATAGTGTGTTTTCCATTCTCAAATTTTTCTCTGGTTTTATATATACATACACACCCCCT	717
QY	736	TCATGTGGCCTACGTC-----CTGCCTCGGAATCAGGAGCATCTACCCCA	780
DB	718	ACTTACCTACCTAATCTACATACACAGTCATGTCCTCACAGTCAGGAAGCATGTACGTTT	777
QY	781	GGTACATCATGAATCGCGCAGCAGTTTGACATTATTGCTGTTTTCAGCTTTTAAGGTTGTT	840
DB	778	GGTACATCCTTGAAATTCAGTAGCAGTTTATTAACTGCTCTTCTCGGTTTAAGATTGTT	837
QY	841	GTCGTTTGTGTTTGTGATTATGTT	863
DB	838	GTGCTTTTGTGTTTAAATTATTI	860

RESULT 15
AC022200/c
LOCUS
DEFINITION Homo sapiens chromosome 8 clone RP11-262017 map 8, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
AC022200 GI:13518207
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavskiy,L., Bouckgalter,B., Brown,A., Buckett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Hesford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marcuis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,B., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo.A., Wu,X., Wyman,D., Ye.W.J.,
Zimmer,A. and Zody,M.

TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 3, 2001 this sequence version replaced gi:13488035.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5196
Center clone name: 262 O 17
----- Summary Statistics
Sequencing vector: M13; M7815; 63% of reads
Sequencing vector: Plasmid; n/a; 37% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 146959 bases at least Q40
Consensus quality: 147487 bases at least Q30

Consensus quality: 147658 bases at least Q20
Insert size: 145000; agarose-fp
Quality size: 147852; sum-of-contigs
Quality coverage: 8.8 in Q20 bases; agarose-fp
Quality coverage: 8.7 in Q20 b.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

- * 1 40424: contig of 40424 bp in length
- * 40425 40524: gap of 100 bp
- * 40525 48708: contig of 8184 bp in length
- * 48709 48808: gap of 100 bp
- * 48809 73107: contig of 24299 bp in length
- * 73108 73207: gap of 100 bp
- * 73208 148152: contig of 74945 bp in length.

FEATURES

Location/Qualifiers
1..148152
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-262017"
/clone_lib="RPC1-11 Human Male BAC"
1..40424
/note="assembly_fragment"
clone_end:SP6
vector_side:left
40525..48708
/note="assembly_fragment"
48809..73107
/note="assembly_fragment"
73208..148152
/note="assembly_fragment"
clone_end:T7
vector_side:right"

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Query Match 61.9%; Score 540.4; DB 2; Length 148152;
Best Local Similarity 86.9%; Pred. No. 1.3e-122;
Matches 643; Conservative 0; Mismatches 86; Indels 11; Gaps 4;
Qy 127 CAGACTCCAGCATATCTTGAAGAACATGCTGAAGAGGAGGCAATGAAGAGATGTACT 186
Db 133780 CTGGCTCCAGGCATATTTTTCAGGGCATGCTGAAGAGG-----CAAATGAAGATGTACT 133727
Qy 187 GGGAGATGAACAGAGGAAGAAACAAAG-CCATTGTAGCTCCCTGTCAAAGAGGAAG 245
Db 133726 GGGAGATGAACAGAGGAAGAAACAAAGCCCATAGAAATCCCTGTCAAAGAGGAAG 133667
Qy 246 AACCCCTCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAAAGTGGTCAAAATTACATCTG 305
Db 133666 AACTCCCTGAAAAAATCTTGTACGTGGAAGCAGAGAGAAAGTGTTAAAAATTACATCTG 133607
Qy 306 AAATACACAGACTGAGAGAAATGAGAGAGGGGTGAACGATTCATATGTACCTGTGAGCT 365
Db 133606 AAATACACAGGAGAGAGAAATGAGAGAGGGGTGAACAGTCCATGTACCCGTGAGCT 133547
Qy 366 TGGAGGTGAAGAAAGCTCTGGCAGCTAGGTGGGATTTCTTCAGTTCCAACAAAG 425
Db 133546 TGTAAGTGAAGAAAGCTCTGGCAGCTAGGTGGGACTTCTTCAGTTCCAACAAAG 133487
Qy 426 GTCTGTATCTGATAACAAACCTATGTTAACTTGGATAGCTGAAGAGAGAGCTCAAA 485
Db 133486 GTCTGTATCTAACAACAACCTATCATGAACCTGGATAGCTAAGAGAGAGAGCTCAA 133427
Qy 486 GATTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAAGCTGAAA 545
|||||

Db 133426 GATTGGTTTGAATGGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAAGTGA 133367
Qy 546 AGAGGAAGAGCGATTTGGGATTTGTCAAAAGTTCAAGCTGGAACCTGGA---ACCACAGAG 602
133366 GGAGGAAGGACTGATTTGGGATTTATCACTGTTTCAAGTGGAAACAGAGGAAACACAGAGG 133307
Qy 603 ATACAGAGGCAAAAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCCCTGATGAAAAGTTC 662
133306 ATACAGAGGCAAAAGATAGGAAAGAGCCTCGGCTATGGGATTTGCCCTGCTGAAAAGTTC 133247
Qy 663 CTGATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTTTCTTTGGTGCACATATATG 722
133246 ATGCTGCTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTGTTGATTTCTTCTCGGTCA 133187
Qy 723 CCTAAATGCAAGTCAATGTCCTTACGTCCTCGCAATGAGGAGCATGTACCCAGG 782
133186 CCTAAATGCAAGTCAAGTCACTTACCTCTGCTTGAATGAGGAGGACATGTTCCCCAGC 133127
Qy 783 TACATCCATCAAGTGGGAGCAGTGTGATTTGCTTTTCAAGCTTTAAGGTTGTTGT 842
133126 TACATGTGGGAACCTCCAGCAGCAATTTGACTTTATTTGATGTTCCAAAC--TTAAGCTGTTGT 133068
Qy 843 GTTTTGTGTTTGTATTATGT 862
133067 GTTTTGTGTTTGTATTATTT 133048

Search completed: October 4, 2005, 22:06:49
Job time : 4052.85 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.
OM nucleic - nucleic search, using sw model
Run on: October 4, 2005, 22:07:00 ; Search time 730.217 Seconds
(without alignments)
8254.755 Million cell updates/sec

Title: US-09-788-476A-3
Perfect score: 873
Sequence: 1 TGGAGTGGGGTAAACAAGATGGCGACCGAGCGGTGGAGCTCCATAGCTAAAGCTTGC 60

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10J_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	873	100.0	873	9	US-09-788-476A-3
2	873	100.0	894	9	Sequence 3, Appli
3	873	100.0	1154	9	Sequence 1, Appli
4	873	100.0	1154	14	US-09-764-846-90
5	853.4	97.8	891	9	Sequence 90, Appl
6	817.8	93.7	1400	21	US-10-091-483-90K
7	817.8	93.7	1818	21	US-09-822-830A-389
					Sequence 389, App
					Sequence 7827, Ap
					Sequence 2592, Ap

8	733.4	84.0	3147	16	US-10-006-285-419	Sequence 419, App
9	690.8	79.1	2553	10	US-09-814-353-20673	Sequence 20673, A
c 10	595.2	68.2	2553	10	US-09-814-353-20673	Sequence 20673, A
11	498.4	57.1	1022	16	US-10-006-285-233	Sequence 233, App
12	458.4	52.5	488	17	US-10-242-535A-29658	Sequence 29658, A
13	458.4	52.5	488	17	US-10-085-783A-29658	Sequence 29658, A
14	455	52.1	470	17	US-10-242-535A-28856	Sequence 28856, A
15	455	52.1	470	18	US-10-085-783A-28856	Sequence 28856, A
16	431	49.4	470	17	US-10-242-535A-40257	Sequence 40257, A
17	431	49.4	470	18	US-10-085-783A-40257	Sequence 40257, A
18	386	44.2	426	17	US-10-242-535A-40571	Sequence 40571, A
19	386	44.2	426	17	US-10-085-783A-40571	Sequence 40571, A
20	379	43.4	379	17	US-10-242-535A-43248	Sequence 43248, A
21	379	43.4	379	18	US-10-085-783A-43248	Sequence 43248, A
22	336.2	38.5	463	10	US-09-918-995-15429	Sequence 15429, A
23	334	38.3	408	10	US-09-930-213-64	Sequence 64, Appl
c 24	265.4	30.4	5469	9	US-09-764-877-4000	Sequence 4000, Ap
c 25	265.4	30.4	5469	10	US-09-764-891-9371	Sequence 9371, Ap
c 26	265.4	30.4	5469	15	US-10-205-428-817	Sequence 817, App
c 27	265.4	30.4	5469	17	US-10-242-515-4000	Sequence 4000, Ap
c 28	265.4	30.4	9453	9	US-09-764-877-3999	Sequence 3999, App
c 29	265.4	30.4	9453	10	US-09-764-891-9370	Sequence 9370, Ap
c 30	265.4	30.4	9453	15	US-10-205-428-816	Sequence 816, App
c 31	265.4	30.4	9453	17	US-10-242-515-3999	Sequence 3999, Ap
c 32	246.2	28.2	300	16	US-10-006-285-196	Sequence 196, App
33	244	27.9	275	20	US-10-425-115-158622	Sequence 158622, Sequence 531, App
34	211.8	24.3	255	10	US-09-930-213-531	Sequence 737, App
35	206	23.6	343	17	US-10-062-674-737	Sequence 1714, Ap
36	174	19.9	257	17	US-10-062-674-1714	Sequence 378, App
37	154.2	17.7	215	9	US-09-728-445-378	Sequence 378, App
38	154.2	17.7	215	22	US-10-964-549-378	Sequence 15759, A
39	92.2	10.6	463	10	US-09-814-353-15759	Sequence 26486, A
c 40	55	6.3	399	22	US-10-450-763-26486	Sequence 29210, A
41	55	6.3	399	22	US-10-450-763-29210	Sequence 32, Appl
42	54	6.2	5659	15	US-10-172-086-32	Sequence 348, App
43	54	6.2	5659	15	US-10-311-455-348	Sequence 40, Appl
44	54	6.2	5659	15	US-10-240-485-40	Sequence 14, Appl
45	54	6.2	5659	19	US-10-311-507-14	

ALIGNMENTS

RESULT 1
US-09-788-476A-3
; Sequence 3, Application US/09788476A
; Patent No. US20020107190A1
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Ching M. et al.
; TITLE OF INVENTION: NOVEL GENES AND EXPRESSION PRODUCTS THEREFROM
; FILE REFERENCE: 3669-0103P
; CURRENT APPLICATION NUMBER: US/09/788.476A
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: PCR Product
US-09-788-476A-3

Query Match	100.0%	Score 873;	DB 9;	Length 873;
Best Local Similarity	100.0%	Pred. No. 2.7e+232;		
Matches	873;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	TGGAGTGGGGTAAACAAGATGGCGACCGAGCGGTGGAGCTCCATAGCTAAAGCTTGC	60	
Db	1	TGGAGTGGGGTAAACAAGATGGCGACCGAGCGGTGGAGCTCCATAGCTAAAGCTTGC	60	
Qy	61	CGAACTAAGCAAGATGCTTCTGCTGTTTGGAGACCAAGGAATAAGCAAGATCT	120	

Db 61 CGAACTAAAGCAAGATGTCTTGCTCGTGGTGGAGACCAAGGAATAAAGCAAGATCT 120
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAAAATGAAGA 180
Db 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAAAATGAAGA 180
Qy 181 TGTACTGGGAGATGAACAGAGGAAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTC 240
Db 181 TGTACTGGGAGATGAACAGAGGAAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTC 240
Qy 241 GGAAGAACCCCTCGAAACAACTGTGTGATGTGGCAGCAGAGAAAGTGGTGAATAAT 300
Db 241 GGAAGAACCCCTCGAAACAACTGTGTGATGTGGCAGCAGAGAAAGTGGTGAATAAT 300
Qy 301 ATCTGAAATACACAGACTGAGAGATGCAAGAGAGGGCTGAAACGATTCATGTACTGT 360
Db 301 ATCTGAAATACACAGACTGAGAGATGCAAGAGAGGGCTGAAACGATTCATGTACTGT 360
Qy 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTCCAAC 420
Db 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTCCAAC 420
Qy 421 AAAAGGTCTGTCTATCTGATTAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGC 480
Db 421 AAAAGGTCTGTCTATCTGATTAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGC 480
Qy 481 TCAAGATTTGGTTGATGTCTTCAATCTCAGAAAGTCTGAAGATGATGAGAAACT 540
Db 481 TCAAGATTTGGTTGATGTCTTCAATCTCAGAAAGTCTGAAGATGATGAGAAACT 540
Qy 541 GAAAAAGAGCAAGAGCGATTTGGGATTTGCACAAAGTTACGCTGGAACTGGAACCCACA 600
Db 541 GAAAAAGAGCAAGAGCGATTTGGGATTTGCACAAAGTTACGCTGGAACTGGAACCCACA 600
Qy 601 GGATACAGAGCAAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 660
Db 601 GGATACAGAGCAAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 660
Qy 661 TCCTGATACCTTCTGTTCTCCAGTGTTCCTCAATTTCTCCTTCTTCTTGGTCAATATA 720
Db 661 TCCTGATACCTTCTGTTCTCCAGTGTTCCTCAATTTCTCCTTCTTCTTGGTCAATATA 720
Qy 721 TGCTTAATGCACAGTCATGCTAGCTCGCTCGCAATGAGGAGCATGTACCCCA 780
Db 721 TGCTTAATGCACAGTCATGCTAGCTCGCTCGCAATGAGGAGCATGTACCCCA 780
Qy 781 GGTACATCCATGAACCTGCGCAGCAGTTTGACTTATTGCTGTTTCAAGTTGTT 840
Db 781 GGTACATCCATGAACCTGCGCAGCAGTTTGACTTATTGCTGTTTCAAGTTGTT 840
Qy 841 GTGTTTTGTTTTGATATTGTTGCTTTGTTAAT 873
Db 841 GTGTTTTGTTTTGATATTGTTGCTTTGTTAAT 873

RESULT 2

US-09-788-476A-1
; Sequence 1, Application US/09788476A
; Patent No. US20020107190A1
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Ching M. et al.
; TITLE OF INVENTION: NOVEL GENES AND EXPRESSION PRODUCTS THEREFROM
; FILE REFERENCE: 3669-0103P
; CURRENT APPLICATION NUMBER: US/09/788,476A
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 894
; TYPE: DNA
; ORGANISM: human
US-09-788-476A-1

Query Match 100.0%; Score 873; DB 9; Length 894;
Best Local Similarity 100.0%; Pred. No. 2,7e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGGAGTGAAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 60
Db 6 TGGAGTGAAGGGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 65
Qy 61 CGAACTAAAGCAAGAAATGTCTTCTGCTGGTGGAGACCAAGGGAAATAAGCAAGATCT 120
Db 66 CGAACTAAAGCAAGAAATGTCTTCTGCTGGTGGAGACCAAGGGAAATAAGCAAGATCT 125
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAAAATGAAGA 180
Db 126 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAAAATGAAGA 185
Qy 181 TGTACTGGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTC 240
Db 186 TGTACTGGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTC 245
Qy 241 GGAAGAACCCCTCGAAACAACTGTGTGATGTGGCAGCAGAGAAAGTGGTGAATAAT 300
Db 246 GGAAGAACCCCTCGAAACAACTGTGTGATGTGGCAGCAGAGAAAGTGGTGAATAAT 305
Qy 301 ATCTGAAATACACAGACTGAGAGATGCAAGAGAGGGCTGAAACGATTCATGTACTGT 360
Db 306 ATCTGAAATACACAGACTGAGAGATGCAAGAGAGGGCTGAAACGATTCATGTACTGT 365
Qy 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTCCAAC 420
Db 366 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTCCAAC 425
Qy 421 AAAAGGTCTGTCTATCTGATTAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGC 480
Db 426 AAAAGGTCTGTCTATCTGATTAACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAAGC 485
Qy 481 TCAAGATTTGGTTGATGTCTTCAATCTCAGAAAGTCTGAAGATGATGAGAAACT 540
Db 486 TCAAGATTTGGTTGATGTCTTCAATCTCAGAAAGTCTGAAGATGATGAGAAACT 545
Qy 541 GAAAAAGAGCAAGAGCGATTTGGGATTTGCACAAAGTTACGCTGGAACTGGAACCCACA 600
Db 546 GAAAAAGAGCAAGAGCGATTTGGGATTTGCACAAAGTTACGCTGGAACTGGAACCCACA 605
Qy 601 GGATACAGAGCAAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 660
Db 606 GGATACAGAGCAAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCCCTGATGAAAGT 665
Qy 661 TCCTGATACCTTCTGTTCTCCAGTGTTCCTCAATTTCTCCTTCTTCTTGGTCAATATA 720
Db 666 TCCTGATACCTTCTGTTCTCCAGTGTTCCTCAATTTCTCCTTCTTCTTGGTCAATATA 725
Qy 721 TGCTTAATGCACAGTCATGCTAGCTCGCTCGCAATGAGGAGCATGTACCCCA 780
Db 726 TGCTTAATGCACAGTCATGCTAGCTCGCTCGCAATGAGGAGCATGTACCCCA 785
Qy 781 GGTACATCCATGAACCTGCGCAGCAGTTTGACTTATTGCTGTTTCAAGTTGTT 840
Db 786 GGTACATCCATGAACCTGCGCAGCAGTTTGACTTATTGCTGTTTCAAGTTGTT 845
Qy 841 GTGTTTTGTTTTGATATTGTTGCTTTGTTAAT 873
Db 846 GTGTTTTGTTTTGATATTGTTGCTTTGTTAAT 878

RESULT 3

US-09-764-846-90
; Sequence 90, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT212

```
; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
; LENGTH: 1154
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1149)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-846-90

Query Match      100.0%; Score 873; DB 9; Length 1154;
Best Local Similarity 100.0%; Pred. No. 3.1e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGTGAGGGGTAAACAAGATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 60
DB 182 TGGAGTGAGGGGTAAACAAGATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 241
QY 61 CGAATTAACCAAGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
DB 242 CGAATTAACCAAGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
QY 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAATGAAGAAGA 180
DB 302 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAATGAAGAAGA 361
QY 181 TGACTGGAGAGTAAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 420
DB 422 GGAAGAACCCCTGMAAAACTGTTGATGTGGCAGCAGAGAAACAAAGAGTGGTGAAGAAATTAC 481
QY 301 ATCTGAATACCAAGACTGAGAGATGCAAGAGAGGCTGGAACGATTCATGTACCTGT 360
DB 482 ATCTGAATACCAAGACTGAGAGATGCAAGAGAGGCTGGAACGATTCATGTACCTGT 541
QY 361 GAGCTTTGGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 420
DB 542 GAGCTTTGGAGAGTAAGAAAGCTCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 601
QY 421 AAAAGGTCTGTCATCTGATTAACAAACCTATGTTGTTAACTTGGATAAGCTGAAGNAAGAGC 480
DB 602 AAAAGGTCTGTCATCTGATTAACAAACCTATGTTGTTAACTTGGATAAGCTGAAGNAAGAGC 661
QY 481 TCAAGATTTGGTTTGAATGCTCTCAATCTCCAGAAAGCTCTGAAGATCATGAGAACT 540
DB 662 TCAAGATTTGGTTTGAATGCTCTCAATCTCCAGAAAGCTCTGAAGATCATGAGAACT 721
QY 541 GAAAAGAGGAGGAGGAGGATTTGGGATTTGTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAG 600
DB 722 GAAAAGAGGAGGAGGAGGATTTGGGATTTGTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAG 781

; CURRENT APPLICATION NUMBER: US/09/764,846
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
; LENGTH: 1154
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1149)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-846-90

Query Match      100.0%; Score 873; DB 9; Length 1154;
Best Local Similarity 100.0%; Pred. No. 3.1e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGTGAGGGGTAAACAAGATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 60
DB 182 TGGAGTGAGGGGTAAACAAGATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGC 241
QY 61 CGAATTAACCAAGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
DB 242 CGAATTAACCAAGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
QY 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAATGAAGAAGA 180
DB 302 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAATGAAGAAGA 361
QY 181 TGACTGGAGAGTAAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 420
DB 422 GGAAGAACCCCTGMAAAACTGTTGATGTGGCAGCAGAGAAACAAAGAGTGGTGAAGAAATTAC 481
QY 301 ATCTGAATACCAAGACTGAGAGATGCAAGAGAGGCTGGAACGATTCATGTACCTGT 360
DB 482 ATCTGAATACCAAGACTGAGAGATGCAAGAGAGGCTGGAACGATTCATGTACCTGT 541
QY 361 GAGCTTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 420
DB 542 GAGCTTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 601
QY 421 AAAAGGTCTGTCATCTGATTAACAAACCTATGTTTAACTTGGATAAGCTGAAGNAAGAGC 480
DB 602 AAAAGGTCTGTCATCTGATTAACAAACCTATGTTTAACTTGGATAAGCTGAAGNAAGAGC 661
QY 481 TCAAGATTTGGTTTGAATGCTCTCAATCTCCAGAAAGCTCTGAAGATCATGAGAACT 540
DB 662 TCAAGATTTGGTTTGAATGCTCTCAATCTCCAGAAAGCTCTGAAGATCATGAGAACT 721
QY 541 GAAAAGAGGAGGAGGAGGATTTGGGATTTGTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAG 600
DB 722 GAAAAGAGGAGGAGGAGGATTTGGGATTTGTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAAG 781
QY 601 GGATACAGAGCAAGAAAGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 782 GGATACAGAGCAAGAAAGAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
QY 661 TCCTGATACCTTCTGTTCTCCAGTGTTTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
DB 842 TCCTGATACCTTCTGTTCTCCAGTGTTTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 901
QY 721 TGCTTAATGACAGTCAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 902 TGCTTAATGACAGTCAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 961
QY 781 GGTACATCCATGACCTGGGAGGAGGATTTGCACTTATTTGCTGTTTCACTTTTAAAGTTGTT 840
```


173	Qy	GAAGAAGATGTACTGGAGATGAAACACAGAGGAAGAGAAACAAAGCCCATTTAGCTCCCT	2332
673	Db	GAAGAAGATGTACTGGAGATGAAACACAGAGGAAGAGAAACAAAGCCCATTTAGCTCCCT	732
233	Qy	GTCAAAGAGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAGAGTGGTG	292
733	Db	GTCAAAGAGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAGAAGAGTGGTG	792
293	Qy	AAAATTACATCTGAAATACACAGACTGAGAGAAATGCAGAAAGAGGGCTGAACGATTCAT	352
793	Db	AAAATTACATCTGAAATACACAGACTGAGAGAAATGCAGAAAGAGGGCTGAACGATTCAT	852
353	Qy	GTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCA	412
853	Db	GTACCTGTGAGCTTGGAGAGTAAGAAAGTGTCTCGGCAGCTAGGTTTGGGATTTCTTCA	912
413	Qy	GTTCACACAAAAGGCTGTGCATCTGATAACAAACCTATGGTTAACTTGGATTAAGCTGAAG	472
913	Db	GTTCACACAAAAGGCTGTGCATCTGATAACAAACCTATGGTTAACTTGGATTAAGCTGAAG	972
473	Qy	GAAAGAGCTCAAAAGATTTGGTTTGGAAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGAT	532
973	Db	GAAAGAGCTCAAAAGATTTGGTTTGGAAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGAT	1032
533	Qy	GAGAAACTGAAAAAGAGGAAGGAGCGATTTGGGATTTGTCACAAGTTTCAGCTGGAACTGGA	592
1033	Db	GAGAAACTGAAAAAGAGGAAGGAGCGATTTGGGATTTGTCACAAGTTTCAGCTGGAACTGGA	1092
593	Qy	ACACAGAGGATACAGAGGCCAAAGAGAGGAAGAGCAGAGCGCTTTGGGATTCCTCGA	652
1093	Db	ACCAAGAGGATACAGAGGCCAAAGAGAGGAAGAGCAGAGCGCTTTGGGATTCCTCGA	1152
653	Qy	TGAAAAAGTCTCTGATACCTTTCTGTTCTCCAGTGTFTTTCCATTTCTCTCCTTCTCTTGGT	712
1153	Db	TGAAAAAGTCTCTGATACCTTTCTGTTCTCCAGTGTFTTTCCATTTCTCTCCTTCTCTTGGT	1212
713	Qy	CACATATATGCTTAAATGCACAGTCATGTGCTACGTCCTGCTCGCAATGAGGGAGCAT	772
1213	Db	CACATATATGCTTAAATGCACAGTCATGTGCTACGTCCTGCTCGCAATGAGGGAGCAT	1272
773	Qy	GTACCCAGGTACATCCATGAATCGCGCAGCAGTTTGCATTATGTGTTGTTTTCAGCTTTA	832
1273	Db	GTACCCAGGTACATCCATGAATCGCGCAGCAGTTTGCATTATGTGTTGTTTTCAGCTTTA	1332
833	Qy	AGGTTGTTGTTTTTGTGTTTGTGATTAATGTTGCTGTTAAT	873
1333	Db	AGGTTGTTGTTTTTGTGTTTGTGATTAATGTTGCTGTTAAT	1373

```

RESULT 7
US-10-956-157-2592
; Sequence 2592, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2592
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-2592

```

Query Match 93.7%; Score 817.8; DB 21; Length 1818;
Best Local Similarity 99.8%; Pred. NO. 9.5e-217;
Matches 819; Conservative 0; Mismatches 2; Indels 0;

Qy	53	AGCTTCCGAACTAAAGCAAGAATGCTGTGCTCGTGGTTTGCAGACCAAGGCAATAAG	112
Db	971	AAACTTCCGAACTAAAGCAAGAATGCTGTGCTCGTGGTTTGCAGACCAAGGGAATAAG	1030
Qy	113	CAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAATGCTGAAGAGGAGGCAAAAT	172
Db	1031	CAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAAATGCTGAAGAGGAGGCAAAAT	1090
Qy	173	GAGGAAGATGTACTGGGAGATGAAAACAGAGGAAGAGAAAACAAAGCCCATTTAGCTCCCT	232
Db	1091	GAGGAAGATGTACTGGGAGATGAAAACAGAGGAAGAGAAAACAAAGCCCATTTAGCTCCCT	1150
Qy	233	GTCAAAGAGGAAGAACCCCTCGAAAAAATCTTGTATGTGCGACGACAGAGAAGAATGGTG	292
Db	1151	GTCAAAGAGGAAGAACCCCTCGAAAAAATCTTGTATGTGCGACGACAGAGAAGAATGGTG	1210
Qy	293	AAAATTTACATCTGAATAACACAGACTGAGAGAAATGCAGAGAGGGCTGAACGATCAAT	352
Db	1211	AAAATTTACATCTGAATAACACAGACTGAGAGAAATGCAGAGAGGGCTGAACGATCAAT	1270
Qy	353	GTACCTGTGAGCTTGGAGAGTAAGAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCA	412
Db	1271	GTACCTGTGAGCTTGGAGAGTAAGAAGTTGCTCGGCAGCTAGGTTTGGGATTTCTTCA	1330
Qy	413	GTTCACAAACAAAGGTCTGTCATCTGTATAACAAACCTATGTTAACTTGGATTAAGCTGAAG	472
Db	1331	GTTCACAAACAAAGGTCTGTCATCTGTATAACAAACCTATGTTAACTTGGATTAAGCTGAAG	1390
Qy	473	GAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGAT	532
Db	1391	GAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGAT	1450
Qy	533	GAGAACTGAAAAGAGGAGGAGCGATTTGGGATTTGTCAACAGTTGACGCTGGAACTGGA	592
Db	1451	GAGAACTGAAAAGAGGAGGAGCGATTTGGGATTTGTCAACAGTTGACGCTGGAACTGGA	1510
Qy	593	ACCACAGAGTACAGAGGCAAGAGAGAAAGAGCAGAGCGCTTTGGGATTCGCTCGA	652
Db	1511	ACCACAGAGTACAGAGGCAAGAGAGAAAGAGCAGAGCGCTTTGGGATTCGCTCGA	1570
Qy	653	TGAAAAGTTCTTGATCTTTCTGTTCTCCAGTGTGTTTTCCATTTCTCTCTTCTTTGGT	712
Db	1571	TGAAAAGTTCTTGATCTTTCTGTTCTCCAGTGTGTTTTCCATTTCTCTCTTCTTTGGT	1630
Qy	713	CACATATATGCTTAAATGCAAGTCATGTGCGCTACGTCCTCGCAATGAGGGAGCAT	772
Db	1631	CACATATATGCTTAAATGCAAGTCATGTGCGCTACGTCCTCGCAATGAGGGAGCAT	1690
Qy	773	GTACCCAGGTACATCCATGAACTGCGGACAGCAGTTTGACTTATTTGCTGTTTCAGCTTTA	832
Db	1691	GTACCCAGGTACATCCATGAACTGCGGACAGCAGTTTGACTTATTTGCTGTTTCAGCTTTA	1750
Qy	833	AGGTTGTTGTTTTGTTTTGTTTTGATTAATGTTGCTGTTAAAT	873
Db	1751	AGGTTGTTGTTTTGTTTTGTTTTGATTAATGTTGCTGTTAAAT	1791

```

RESULT 8
US-10-006-285-419
; Sequence 419, Application US/10006285
; Publication No. US20030165854A1
; GENERAL INFORMATION:
; APPLICANT: Mary Jane Cunningham
; APPLICANT: Matthew R. Kaser
; TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS
; FILE REFERENCE: PA-0039 US
; CURRENT APPLICATION NUMBER: US/10/006,285
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 514
; SOFTWARE: PERL Program
; SEQ ID NO 419
; LENGTH: 3147

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165854A1 235169.27
US-10-006-285-419

Query Match      84.0%; Score 733.4; DB 16; Length 3147;
Best Local Similarity 99.9%; Pred. No. 4.1e-193;
Matches 734; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 52 AAAGCTTGGCGAACTTAAGCAAGATGTCCTGCTGCTGGTTGGAGACCAAGGAATAAA 111
Db 97 AAAGCTTGGCGAACTTAAGCAAGATGTCCTGCTGCTGGTTGGAGACCAAGGAATAAA 156
QY 112 GCAAGATCTTATCCACAGACTCCAGCATATCTTGAAGCAATCTGTAAGAGAGGCAAA 171
Db 157 GCAAGATCTTATCCACAGACTCCAGCATATCTTGAAGCAATCTGTAAGAGAGGCAAA 216
QY 172 TGAAGAAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCC 231
Db 217 TGAAGAAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCC 276
QY 232 TGTCAAAGAGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAAAGTGGT 291
Db 277 TGTCAAAGAGGAAGAACCCCTGAAAAAATCTGTTGATGTGGCAGCAGAGAAAGTGGT 336
QY 292 GAAATTTACATCTGAAATACACACAGACTGAGAGATGCAGAGAGGGCTGAACCATCAA 351
Db 337 GAAATTTACATCTGAAATACACACAGACTGAGAGATGCAGAGAGGGCTGAACCATCAA 396
QY 352 TGTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTC 411
Db 397 TGTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTC 456
QY 412 AGTTCCAAACAAAGGCTCTCATCTGATAACAAACCTATAGTTAACTTGGATAAGCTGAA 471
Db 457 AGTTCCAAACAAAGGCTCTCATCTGATAACAAACCTATAGTTAACTTGGATAAGCTGAA 516
QY 472 GGAAGAGCTCAAGATTTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGA 531
Db 517 GGAAGAGCTCAAGATTTGGTTTGAATGTCTTCAATCTCCAGAAAGTCTGAAGATGA 576
QY 532 TGAGAACTTGAAAGAGAGGAGCGATTGGGATTTGTCAAGTTTCAGCTCGAACTGG 591
Db 577 TGAGAACTTGAAAGAGAGGAGCGATTGGGATTTGTCAAGTTTCAGCTCGAACTGG 636
QY 592 AACCACAGAGGATACAGAGGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 651
Db 637 AACCACAGAGGATACAGAGGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 696
QY 652 ATGAAAAGTTCTGTATCTTCTGTTCCAGTGTTTCCATTTCTCTCTCTCTCTCTCTGG 711
Db 697 ATGAAAAGTTCTGTATCTTCTGTTCCAGTGTTTCCATTTCTCTCTCTCTCTCTCTGG 756
QY 712 TCACATATATGCTTAATGCACAGTCATGCTGCTACGTCCTGCTCGCAATGAGGGAGCA 771
Db 757 TCACATATATGCTTAATGCACAGTCATGCTGCTACGTCCTGCTCGCAATGAGGGAGCA 816
QY 772 TGTACCCAGGTACA 786
Db 817 TGTACCCAGGTACA 831

RESULT 9
US-09-814-353-20673
; Sequence 20673, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS AND METHODS FOR

```



```
Qy 753 GCCTGCAATCAGGAGCATGTACCCAGGTACA 786
Db 1375 GCCTGCAATGAGGAGCATGTACCCAGGTCA 1408

RESULT 10
US-09-814-353-20673/c
; Sequence 20673, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20673
; LENGTH: 2553
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-20673

Query Match 68.2%; Score 595.2; DB 10; Length 2553;
Best Local Similarity 96.9%; Pred. No. 1.1e-154;
Matches 628; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

Qy 202 GGAAGAGAAACAAGCCCAATTGAGTCCCTGTCAAAGAGGAAGAACCCCTG-AAAAA 260
Db 695 GGAAGAGAAACAAGCCCAATTGAGTCCCTGTCAAAGAGGAAGAACCCCTGAAAAA 636

Qy 261 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATACATCTGAATACCAAGACTG 320
Db 635 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATACATCTGAATACCAAGACTG 576

Qy 321 AGAGAATGCAAGAGGAGGCTGAACGATTCAATGTACCTGTGAGCTTGGAGAGTAAGAAAG 380
Db 575 AGAGAATGCAAGAGGAGGCTGAACGATTCAATGTACCTGTGAGCTTGGAGAGTAAGAAAG 516

Qy 381 CTGTCGGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAAGGCTGTCACTCGATA 440
Db 515 CTGTCGGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAAGGCTGTCACTCGATA 456

Qy 441 ACAAACTATGGTTAACTTTGATTAAGCTGAAGGAAAGAGCTCAAGATTGGTTTGAATG 500
Db 455 ACAAACTATGGTTAACTTTGATTAAGCTGAAGGAAAGAGCTCAAGATTGGTTTGAATG 396

Qy 501 TCTCTCAATCTCAGAAAGTCTGAAGATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 560
Db 395 TCTCTCAATCTCAGAAAGTCTGAAGATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 336

Qy 561 TTGGGATGTCAAGATTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 620
Db 335 TTGGGATGTCAAGATTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 276

Qy 621 GGAAGAGCAGAGCGCTTGGGATTTGCGATGAAAGTTCTGATGATACTTTCTGTTCTC 680
Db 621 GGAAGAGCAGAGCGCTTGGGATTTGCGATGAAAGTTCTGATGATACTTTCTGTTCTC 680

Db 275 GGAAGAGCAGAGCGCTTGGGATTTGCGATGAAAGTTCTGATGATACTTTCTGTTCTC 216
Qy 681 CAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 740
Db 215 CAGTGTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 156
Qy 741 TGCTACGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 800
Db 155 TGCTACGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 96
Qy 801 CAGCAGTTCAGTTCATTTGCTGTTTTCAGCTTTTAAGGTTGTTGTTGTTT 848
Db 95 GCACAGTTCAGTTCATTTGCTGTTTTCAGCTTTTAAGGTTGTTGTTGTTT 49

RESULT 11
US-10-006-285-233
; Sequence 233, Application US/10006285
; Publication No. US20030165854A1
; GENERAL INFORMATION:
; APPLICANT: Mary Jane Cunningham
; APPLICANT: Matthew R. Kaser
; TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS
; FILE REFERENCE: PA-0039 US
; CURRENT APPLICATION NUMBER: US/10/006,285
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 514
; SOFTWARE: PERL Program
; SEQ ID NO 233
; LENGTH: 1022
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165854A1 207354_Rn.3
; NAME/KEY: unsure
; LOCATION: 772-845
; OTHER INFORMATION: a, t, c, g, or other
US-10-006-285-233

Query Match 57.1%; Score 498.4; DB 16; Length 1022;
Best Local Similarity 87.3%; Pred. No. 6.4e-128;
Matches 580; Conservative 0; Mismatches 81; Indels 3; Gaps 3;

Qy 52 AAAGCTTTCGCGAATTAAGCAAGAAATCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 111
Db 104 AAAGCTTTCGCGAATTAAGCAAGAAATCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 163

Qy 112 GCAAGATCTTATCCACAGACT-CCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCGAA 170
Db 164 ACAAGATCTTATCAATAGGCTACCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCGAA 223

Qy 171 ATGAAGAAGATGTACTGGGAGATGAACAGAGGAAGAAACAAACCCATTCAGCTCC 230
Db 224 ATGAAGAAGATGTACTGGGAGATGAACAGAGGAAGAAACCAAGCCTATAGACTGC 283

Qy 231 CTGTCAAAGAGGAAGAACCCCTGAAAACAACTGTTGATGTGGCAGCAGAGAGAAAGTGG 290
Db 284 CTGTCAAAGAGGAAGAACCCCTGAAAACAACTGTTGATATGGCATCAGAAAAGAGTGG 343

Qy 291 TGAATAATCATCTGAAATACCAAGCTGAGAGAAATGCAAGAGAGGCTGAAAGATTCA 350
Db 344 TAAATAATCATCTGGAATACCTCAAACTGAGAGAAATGCAAGAGAGGCTGAAAGATTCA 403

Qy 351 ATGTACCTGTGAGCTTGGAGAGTAAGAAAGTCTGCTG-GGCAGCTAGGTTTGGATTTCT 409
Db 404 ATGTGCTGTGAAGCTTGGAGAGTAAGAAAGTCTGCTGCGGAGCAGAGGTTTGGAAATTTCT 463

Qy 410 TCAGTTCCAAACAAAGGTTGTTGATCTGATAACAAACCTATGTTAACTTGGATAAGCTG 469
Db 464 TCAGTTCCAAACAAAGGTTTATCATCTGACACCAAGCAATGGTTAACTTGGATAACTA 523

Qy 470 AAGGAAAGAGCTCAAGAGTTTGGTTTGAATGTCTTTCAATCTCCAGAAAGTCTGAAGAT 529
```



```
Db 305 ACATCTGAAATACACAGACTGAGAGAAATGCAGAAGAGGCTGAACGATTCAATGTACCT 364
Qy 359 GTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCGACGTAGGTTTGGGATTTCTTCAGTTCCA 418
Db 365 GTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCGACGTAGGTTTGGGATTTCTTCAGTTCCA 424
Qy 419 ACAAAGGTCGTCTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAGA 478
Db 425 ACAAAGGTCGTCTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAGA 484
Qy 479 GCTC 482
Db 485 GCTC 488

RESULT 14
US-10-242-535A-28856
; Sequence 28856, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 28856
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Human
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (23)..(23)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-28856

Query Match 52.1%; Score 455; DB 17; Length 470;
Best Local Similarity 99.8%; Pred. No. 5.2e-116;
Matches 466; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TGGAGTGAGGGGTAACAAG-ATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTG 59
Db 4 TGGAGTGAGGGGTAACAAGNATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTG 63
Qy 60 CCGAACTAAAGCAAGAATGTCTTGTCTGTTGGAGACCAAGGGAATAAGCAAGATC 119
Db 64 CCGAACTAAAGCAAGAATGTCTTGTCTGTTGGAGACCAAGGGAATAAGCAAGATC 123
Qy 120 TTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAATGAAGAG 179
Db 124 TTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAATGAAGAG 183
Qy 180 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 239
Db 184 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 243
Qy 240 AGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATTA 299
Db 244 AGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATTA 303
Qy 300 CATCTGAAATACCAAGAGTGAAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 359
Db 184 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 243
Qy 240 AGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATTA 299
Db 244 AGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATTA 303
Qy 300 CATCTGAAATACCAAGAGTGAAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 359
```

```
Db 304 CATCTGAAATACCAAGAGTGAAGAGGCTGAACGATTCAATGTACCTG 363
Qy 360 TGAGCTTGGAGAGTAAGAAAGCTGCTCGGCGACGTAGGTTTGGGATTTCTTCAGTTCCA 419
Db 364 TGAGCTTGGAGAGTAAGAAAGCTGCTCGGCGACGTAGGTTTGGGATTTCTTCAGTTCCA 423
Qy 420 CAAAAGGTCGTCTCATCTGATAACAAACCTATGTTAACTTGGATAAG 466
Db 424 CAAAAGGTCGTCTCATCTGATAACAAACCTATGTTAACTTGGATAAG 470

RESULT 15
US-10-085-783A-28856
; Sequence 28856, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 28856
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Human
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (23)..(23)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-28856

Query Match 52.1%; Score 455; DB 18; Length 470;
Best Local Similarity 99.8%; Pred. No. 5.2e-116;
Matches 466; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 TGGAGTGAGGGGTAACAAG-ATGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTG 59
Db 4 TGGAGTGAGGGGTAACAAGNATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTG 63
Qy 60 CCGAACTAAAGCAAGAATGTCTTGTCTGTTGGAGACCAAGGGAATAAGCAAGATC 119
Db 64 CCGAACTAAAGCAAGAATGTCTTGTCTGTTGGAGACCAAGGGAATAAGCAAGATC 123
Qy 120 TTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAATGAAGAG 179
Db 124 TTATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGAGGCAATGAAGAG 183
Qy 180 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 239
Db 184 ATGTACTGGGAGATGAACAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 243
Qy 240 AGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATTA 299
Db 244 AGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGTGGTGAATAATTA 303
Qy 300 CATCTGAAATACCAAGAGTGAAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 359
Db 304 CATCTGAAATACCAAGAGTGAAGAGGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAG 363
Qy 360 TGAGCTTGGAGAGTAAGAAAGCTGCTCGGCGACGTAGGTTTGGGATTTCTTCAGTTCCA 419
Db 364 TGAGCTTGGAGAGTAAGAAAGCTGCTCGGCGACGTAGGTTTGGGATTTCTTCAGTTCCA 423
Qy 420 CAAAAGGTCGTCTCATCTGATAACAAACCTATGTTAACTTGGATAAG 466
```

Db 424 CAAAAGGCTGTGTCATCTGTGATAACAAACCTATGGTTAACTTGGATAAG 470

Search completed: October 5, 2005, 05:55:42
Job time : 735.217 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:21:02 ; Search time 176.873 Seconds
(without alignments)
8076.260 Million cell updates/sec

Title: US-09-788-476A-3
Perfect score: 873
Sequence: 1 tggagtgagggttaacaaga.....tgattatgtgtgtgtaaat 873

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	647.4	74.2	22303	4	US-09-949-016-12616
C 2	647.4	74.2	22303	4	US-09-949-016-12964
C 3	525	60.1	553	4	US-09-621-976-3627
C 4	465.6	53.3	471	4	US-09-513-999C-736
C 5	374.2	42.9	405	4	US-09-621-976-18639
C 6	331.6	38.0	383	4	US-09-621-976-18638
C 7	241	27.6	337	4	US-09-621-976-18236
C 8	79.4	9.1	7218	1	US-08-232-463-14
C 9	75	8.6	601	4	US-09-949-016-35088
C 10	75	8.6	601	4	US-09-949-016-42456
C 11	74.6	8.5	601	4	US-09-949-016-35089
C 12	74.6	8.5	601	4	US-09-949-016-42457
C 13	53.6	6.1	3211	2	US-08-574-959A-8
C 14	53.6	6.1	3211	3	US-08-574-959A-6
C 15	53.6	6.1	3901	2	US-09-357-014-8
C 16	53.6	6.1	3901	3	US-09-357-014-6
C 17	52.8	6.0	64309	4	US-09-949-016-14581
C 18	52	6.0	92227	4	US-09-949-016-11929
C 19	52	6.0	92232	4	US-09-949-016-15421
C 20	47.6	5.5	929	4	US-09-671-317-14
C 21	46.6	5.3	2754	4	US-09-248-796A-4857
C 22	45.6	5.2	601	4	US-09-949-016-159256
C 23	45.6	5.2	4756	4	US-09-949-016-4455
C 24	45.6	5.2	53737	4	US-09-949-016-16197
C 25	45.4	5.2	7044	4	US-09-949-016-14113
C 26	45.2	5.2	157032	4	US-09-949-016-16502
C 27	45	5.2	601	4	US-09-949-016-90371

ALIGNMENTS

RESULT 1

US-09-949-016-12616/c
; Sequence 12616, Application US/09949016
; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CU001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12616

; LENGTH: 22303

; TYPE: DNA

; ORGANISM: Human

US-09-949-016-12616

Query Match 74.2%; Score 647.4; DB 4; Length 22303;

Best Local Similarity 90.4%; Pred. No. 5.3e-189;

Matches 782; Conservative 0; Mismatches 46; Indels 37; Gaps 7;

Qy	7	GAGGGGTAACAAGATGCTTCTGCTCGTGGTGGAGCGGTGGAGCTCCATAAGCTAAAGCTTGCGGAAC	66
Db	6441	GAGGGGTAACAAGATGCTTCTGCTCGTGGTGGAGCGGTGGAGCTCCATAAGCTTGCGGAAC	6382
Qy	67	AAAGCAAGATGCTTCTGCTCGTGGTGGAGCGGTGGAGCTCCATAAGCTTGCGGAAC	126
Db	6381	AAAGCAAGATGCTTCTGCTCGTGGTGGAGCGGTGGAGCTCCATAAGCTTGCGGAAC	6323
Qy	127	CAGACTCCAGGCATATCTTGAAGCAATCTGTAAGAGAGGCAATGAAGAGATGACT	186
Db	6322	CAGACTCCAGGCATATCTTGAAGCAATCTGTAAGAGAGGCAATGAAGAGATGACT	6266
Qy	187	GGGAGATGAACAGAGAGAGCAAAAGCCATTTGAGCTCCCTGTCGAAGAGGAGA	246
Db	6265	AGGAGATGAACAGAGAGAGCAAAAGCCATTTGAGCTCCCTGTCGAAGAGGAGA	6219
Qy	247	ACCCCTTG-AAAACCTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATACATCTG	305
Db	6218	ACCCCTTG-AAAACCTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATACATCTG	6159
Qy	306	AAATACCACAGACTGAGAGAAATGCAGAGAGGGCTGAACCGATTCAATGTACCTGTGAGCT	365

```
Db 6158 AATACCAAGGCTGAGAGAAATGAGAGAGAGGCGGCAAACTTCAGTGATCTGTGAGCT 6099
QY 366 TGGAGAGTAAGAAAGCTGCTGGGACAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAG 425
Db 6098 TGGAGAGTAAGAAAGCTGCTCAGGACAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAG - 6040
QY 426 GTCTGTCATCTGATAACAAACCTATGTTTAACTTGGATAGCTGAGGAAAGAGCTCAAA 485
Db 6039 GTCTGTCATCTG--AACACACCTATGTTTAACTTGGATAGCTGAGGAAAGAGCTCAAA 5982
QY 486 GATTGTGTTTGAATGCTCTTCAATCTCCAGAAAGCTGAGAGATGATGAGAAACTGAAA 545
Db 5981 GATTGTGTTTGAATGCTCTTCAATCTCCAGAAAGCTGAGAGATGATGAGAAACTGAAA 5922
QY 546 AGAGGAAGAGCGGATTTGGGATTTGTCAAAAGTTTCAGCTGGAACTGGAAACACACAGAGATA 605
Db 5921 AGAGGAAGAGCGGATTTGGGATTTGTCAAAAGTTTCAGCTGGAACTGGAAACACACAGAGATA 5862
QY 606 CAGAGGCAAAAGAGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTCCCTG 665
Db 5861 CAGAGGCAAAAGAGAGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTCCCTG 5802
QY 666 ATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 725
Db 5801 ATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5742
QY 726 AATGCAAGTATGTCCTACGCTTCCTCGCAATGAGGAGAGCATGTATCCCAAGGTAC 785
Db 5741 AA-----TGCAGTCTTGGCTTGCATGAGGAGAGCATGTATCCCAAGGTAA 5698
QY 786 ATCCATGAACCTGGGAGAGAGTTGACTTATGCTGTTTTCAGCTTTAAGTTGTTGTT 845
Db 5697 AACTGTGAACCTGGGAGAGAGTTGACTTATGCTGTTTTCAGCTTTAAGTTGTTGTT 5638
QY 846 TTTGTTTTTGATTATGTTGTTGTT 870
Db 5637 TTTGTTTTTGATTATGTTGTTGTT 5613
```

RESULT 2

```
US-09-949-016-12964/c
; Sequence 12964, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12964
; LENGTH: 22303
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12964
```

```
Query Match 74.2%; Score 647.4; DB 4; Length 22303;
Best Local Similarity 90.4%; Pred No. 5.3e-189;
Matches 782; Conservative 0; Mismatches 46; Indels 37; Gaps 7;
QY 7 GAGGGGTAAACAAGATGGCGGAGAGCGGTGGAGCTCCATAAGCTTAAAGCTTCCGAACT 66
Db 6441 GAGGGGTAAACAAGATGGTAACTGAGATGGTGGAGCTCCATAAGCTTGAAGCTTGTAACT 6382
```

```
QY 67 AAAGCAAGAACTCTTGTCTGCTGTTGGAGACCAAGGGAATAAAACCAAGATCTTATCCA 126
Db 6381 AAAGCAAGAACTCTTGTCTGCTGTTGGAGACCAAGGGAATAAAACCAAGATCTTATCCA 6323
QY 127 CAGACTCCAGGCAATATCTTGAAGAAACATCTGAAAGAGAGGCAAAATGAAAGAAAGATGACT 186
Db 6322 CAGACTCCAGGCAATATCTTGAAGAAACATCTGAAAGAGAGGCAAAAT--GAAGATGACT 6266
QY 187 GGGAGATGAAACAGAGAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTGTAAGAGGAAGA 246
Db 6265 AGGAGATGAAACAGAGAGAAAGAAACAAAGCCCATTTGAGCTCCCTGTGTAAGAGGAAGA 6219
QY 247 ACCCCCTG--AAAGAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATTTACATCTG 305
Db 6218 ACCCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATTTACATCTG 6159
QY 306 AATATCACACAGACTGAGAGAAATGCAGAGAGGCGCTGAACGATTTCAATGTACCTGTGAGCT 365
Db 6158 AATATCACACAGCTGAGAGAAATGCAGAGAGGCGCGCAACATTCAGTGTACCTGTGAGCT 6099
QY 366 TGGAGATGAAAGAAAGCTGCTCGGCGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAG 425
Db 6098 TGGAGATGAAAGAAAGCTGCTCAGGCGAGCTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAG - 6040
QY 426 GTCTGTCATCTGATAACAAACCTATGTTTAACTTGGATAGCTGAGGAAAGAGCTCAAA 485
Db 6039 GTCTGTCATCTG--AACACACCTATGTTTAACTTGGATAGGCGGAAAGAGAGCTCAAA 5982
QY 486 GATTGTGTTTGAATGCTCTTCAATCTCCAGAAAGTCTCAAGATGATGAGAAACTGAAAA 545
Db 5981 GATTGTGTTTGAATGCTCTTCAATCTCCAGAAAGTCTCAAGATGATGAGAAACTGAAAA 5922
QY 546 AGAGGAAGAGCGGATTTGGGATTTGTCAAAAGTTTCAGCTGGAACCTGGAACCAAGAGATA 605
Db 5921 AGAGGAAGAGCGGATTTGGGATTTGTCAAAAGTTTCAGCTGGAACCTGGAACCAAGAGATA 5862
QY 606 CAGAGCAAGAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTCCCTG 665
Db 5861 CAGAGCAAGAGAGAGGAAAGAGCAGAGCGCTTTGGGATTTGCTGATGAAAAGTTCCCTG 5802
QY 666 ATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 725
Db 5801 ATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5742
QY 726 AATGCAAGTATGTCCTACGCTTCCTCGCAATGAGGAGAGCATGTATCCCAAGGTAC 785
Db 5741 AA-----TGCAGTCTTGGCTTGCATGAGGAGAGCATGTATCCCAAGGTAA 5698
QY 786 ATCCATGAACCTGGGAGAGAGTTGACTTATGCTGTTTTCAGCTTTAAGTTGTTGTT 845
Db 5697 AACTGTGAACCTGGGAGAGAGTTGACTTATGCTGTTTTCAGCTTTAAGTTGTTGTT 5638
QY 846 TTTGTTTTTGATTATGTTGTTGTT 870
Db 5637 TTTGTTTTTGATTATGTTGTTGTT 5613
```

RESULT 3

```
US-09-621-976-3627
; Sequence 3627, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 3627
; LENGTH: 553
```


Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:20:27 ; Search time 3485.58 Seconds
(without alignments)
9533.602 Million cell updates/sec

Title: US-09-788-476A-3

Perfect score: 873

Sequence: 1 tggagtggggtaacaaga.....tgattatgttctgttaat 873

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_hic.*
- 4: gb_est3.*
- 5: gb_est4.*
- 6: gb_est5.*
- 7: gb_est6.*
- 8: gb_gss1.*
- 9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848.2	97.2	871	5	BM916484
2	832.4	95.3	1000	7	CO580484
3	830.8	95.2	874	5	BM915580
4	825.2	94.5	877	5	BU849740
5	824.4	94.4	895	5	BU174287
6	824	94.4	1030	4	BM559381
7	820.4	94.0	963	7	CO648068
8	820	93.9	844	3	CF610192
9	820	93.9	910	3	AF161434
10	814.4	93.3	1070	4	BM460786
11	814	93.2	928	5	BU855435
12	811.4	92.9	817	5	BU456775
13	801.6	91.8	843	5	BU456776
14	798.8	91.5	823	5	BU599301
15	798.2	91.4	914	5	BU508603
16	797.4	91.3	910	5	BU157949
17	784.2	89.8	842	5	BQ692014
18	784	89.8	784	3	CR593273
19	783.8	89.8	937	5	BU856660
20	782	89.6	794	7	CN261593
21	782	89.6	1006	5	BQ068156
22	780.8	89.4	795	7	CN261585
23	780	89.3	806	4	BG574651
24	776.2	88.9	883	5	BU146608

25	768.6	88.0	779	5	BM916484
26	762.8	87.4	1105	7	CO580484
27	759	86.9	783	4	BG910161
28	758.8	86.9	1104	5	BM912983
29	758	86.8	914	5	BU902964
30	757	86.7	818	5	BQ227687
31	756.8	86.7	902	5	BQ62276
32	756.6	86.7	783	6	CA775512
33	755.8	86.6	760	4	BM722636
34	755	86.5	788	4	BM462456
35	753.6	86.3	876	5	BU855417
36	750.8	86.0	876	5	BU194990
37	749.4	85.8	783	4	BG533012
38	744.2	85.2	791	6	CB956735
39	742.2	85.0	878	5	BQ213428
c 40	741.2	84.9	785	5	BU632453
41	740.2	84.8	776	4	BI460375
42	738.2	84.6	887	4	BI255433
43	738	84.5	738	7	CF130101
44	738	84.5	819	5	BU596410
c 45	736.4	84.4	753	5	BM979030

ALIGNMENTS

RESULT 1
BM916484
LOCUS BM916484 871 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT 6641858 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5482902
5', mRNA sequence.
ACCESSION BM916484
VERSION BM916484.1 GI:19366863
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 871)
TITLE NIH-MGC http://mgs.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-x@mail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at:
http://image.llnl.gov
Plate: LLCM2009 row: 1 column: 07
High quality sequence stop: 748.
Location/Qualifiers
1. .871
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5482902"
/tissue_types="amelanotic melanoma, cell line"
/lab_host="DHI08 (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

FEATURES

source
1. .871
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5482902"
/tissue_types="amelanotic melanoma, cell line"
/lab_host="DHI08 (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 97.2% Score 848.2 DB 5 Length 871;

Best Local Similarity 99.5%; Pred. No. 4.7e-212;		Matches 850; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY	20	ATGGGACCGAGAGCGTGGAGCTCCATAAGCTTAAGCTTCCCGAATTAAGCAAGAAATGT	79
Db	1	ATGGGACCGAGAGCGTGGAGCTCCATAAGCTTAAGCTTCCCGAATTAAGCAAGAAATGT	60
QY	80	CTTGCTCGTGTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGACTCCAGGCA	139
Db	61	CTTGCTCGTGTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGACTCCAGGCA	120
QY	140	TATCTTTGAAGAACATGCTGAAGAGGCGCAAAATGAAGAGATGTACTGGAGATGAACA	199
Db	121	TATCTTTGAAGAACATGCTGAAGAGGCGCAAAATGAAGAGATGTACTGGAGATGAACA	180
QY	200	GAGGAAGAGAAACAAAGCCATTGAGCTCCCTGTCAAGAGAGAAAGAACCCCTGAAAAA	259
Db	181	GAGGAAGAGAAACAAAGCCATTGAGCTCCCTGTCAAGAGAGAAAGAACCCCTGAAAAA	240
QY	260	ACTGTTGATGTGCGCAGCAGAGAGAAAGTGTGAAATTTACATCTGAAATACCACAGACT	319
Db	241	ACTGTTGATGTGCGCAGCAGAGAGAAAGTGTGAAATTTACATCTGAAATACCACAGACT	300
QY	320	GAGAGAAATGCAGAGAGGCGTGAACGATTCATGCTGTGAGCTTGGAGAGTAAGAAA	379
Db	301	GAGAGAAATGCAGAGAGGCGTGAACGATTCATGCTGTGAGCTTGGAGAGTAAGAAA	360
QY	380	GCTGCTCGGCGAGTAGTGTGGATTTCTTCAGTTTCCAAACAAAGTCTGTCTCATGTAT	439
Db	361	GCTGCTCGGCGAGTAGTGTGGATTTCTTCAGTTTCCAAACAAAGTCTGTCTCATGTAT	420
QY	440	AACAAACCTATGTTAACTTGGATAGCTCAAGGAAAGACTCAAAAGATTTGGTTTGAAT	499
Db	421	AACAAACCTATGTTAACTTGGATAGCTCAAGGAAAGACTCAAAAGATTTGGTTTGAAT	480
QY	500	GTCTCTTCAATCTCCGAAAGTCTGAAGATGATGAGAAACTGAAAAAGAGGAGGCGA	559
Db	481	GTCTCTTCAATCTCCGAAAGTCTGAAGATGATGAGAAACTGAAAAAGAGGAGGCGA	540
QY	560	TTTGGGATTTCAAGATTGAGTGGAACTGGAACCAACAGAGATACAGAGCAAGAAAG	619
Db	541	TTTGGGATTTCAAGATTGAGTGGAACTGGAACCAACAGAGATACAGAGCAAGAAAG	600
QY	620	AGGAAAGAGCAGAGCGCTTGGATGCTGATGAAAGTTTCTGATATCTTCTCTTCT	679
Db	601	AGGAAAGAGCAGAGCGCTTGGATGCTGATGAAAGTTTCTGATATCTTCTCTTCT	660
QY	680	CCAGTGTGTTTCCATTTCTCTCTTCTTCTTGGTCAATATATGCTAAATGCACAGTCAT	739
Db	661	CCAGTGTGTTTCCATTTCTCTCTTCTTGGTCAATATATGCTAAATGCACAGTCAT	720
QY	740	GTGCTACGTCCTGCGCAATGAGGAGGAGCATGTACCCAGAGTACATCCATGAATGCG	799
Db	721	GTGCTACGTCCTGCGCAATGAGGAGGAGCATGTACCCAGAGTACATCCATGAATGCG	780
QY	800	GCAGCAGTTTGACTTATGCTGTTTCAAGCTTAAAGTTGTTGTTTCTTTTTCATTA	859
Db	781	GCAGCAGTTTGACTTATGCTGTTTCAAGCTTAAAGTTGTTGTTTCTTTTTCATTA	840
QY	860	TGTTGCTTGTGTAAT	873
Db	841	TGGTGCTTGTGTAAT	854
RESULT 2			
CO580484			
LOCUS			
DEFINITION			
ILLUMIGEN MCQ 48614 Katze MMTF Macaca mulatta cDNA clone			
TBIUW:18168 57 similar to Bases 522 to 900 highly similar to human			
CIP29 (Hb.410597), mRNA sequence.			
CO580484			
ACCESSION			
VERSION			
CO580484.1 GI:50411638			
KEYWORDS			
EST.			
SOURCE			
ORGANISM			
Macaca mulatta (rhesus monkey)			
Macaca mulatta			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Euthera; Primates; Catarrhini; Cercopitheidae;			
Cercopitheinae; Macaca.			
REFERENCE			
1 (bases 1 to 1000)			
Katze, M.G., Thomas, M., Korth, M., Iadonato, S.P. and Magness, C.L.			
Large-scale Rhesus Macaque cDNA Sequencing			
Unpublished (2003)			
COMMENT			
Contact: C. Magness			
Illumigen Biosciences Inc.			
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA			
Tel: 2063780400			
Fax: 2063780408			
Email: cmagnes@illumigen.com			
Sequenced on 2004.07.02. 775 Q20 bases. Library Preparation: Prof.			
Michael Katze Lab at University of Washington DNA Sequencing:			
Illumigen Biosciences Inc. For further information, see			
http://www.macaque.org			
PCR Primers			
FORWARD: CCTCACTAAAGGGACAAAA			
BACKWARD: CACTATAGGGCAATTTGGTA			
Insert length: 1000 Std Error: 0.00			
Plate: CL000408 row: B column: 10			
Seq primer: CCTCACTAAAGGGACAAAA			
POLYA=Yes.			
FEATURES			
source			
1..1000			
/organism="Macaca mulatta"			
/mol_type="mRNA"			
/strain="Indian"			
/db_xref="taxon:9544"			
/clone="TBIUW:18168"			
/sex="male"			
/tissue_type="testis"			
/dev_stage="juvenile"			
/lab_host="Electromax DH10B"			
/clone_lib="Katze MMTF"			
/note="Vector: pDONR 222; Site 1: BsrG I; Site 2: BsrG I;			
Created from Cloneminer cDNA Library Construction kit			
(catalog #18249-029)"			
ORIGIN			
Query Match 95.3%; Score 832.4; DB 7; Length 1000;			
Best Local Similarity 98.1%; Pred. No. 7e-208;			
Matches 842; Conservative 0; Mismatches 16; Indels 0; Gaps 0;			
QY	16	CAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGCAGAACTAAAGCAAGA	75
Db	5	CAAGATGGCGACCGAAACGGTGGAGCTCCATAAGCTGAAGCTTGTGAACCTAAAGCAAGA	64
QY	76	ATGTCCTTGCTGTTGGTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCA	135
Db	65	ATGTCCTTGCTGTTGGTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCA	124
QY	136	GCATATCTTGAAGAACATGCTGAAGAGGAGCAAAATGAAGAGATGTACTGGGAGATGA	195
Db	125	GCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGATGTACTGGGAGATGA	184
QY	196	AACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTCAAGAGAGAAAGAACCCCTGTA	255
Db	185	AACAGAGGAAGAAACAAAGCCATTGAGCTACCTGTCAAGAGAGAAAGAACCCCTGTA	244
QY	256	AAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTACATCTGAAATACCACA	315
Db	245	AAAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATTACATCTGAAATACCACA	304
QY	316	GACTGAGAGAAATGCAGAGAGGCGTCAACGATTTCAATGTACCTGTGAGCTTTGGAGAGTAA	375
Db	305	GACTGAGAGAAATGCAGAGAGGCGTCAACGATTTCAATGTACCTGTGAGCTTTGGAGAGTAA	364
QY	376	GAAAGCTGCTCGGCGCAGCTAGGTTTGGGATTTCTTCAGTTCCTCAACAAAAAGGTCTGTCATC	435

RESULT 4

BUS49740 877 bp mRNA linear EST 16-OCT-2002
 LOCUS AGENCOURT 10440969 NIH_MGC_109 Homo sapiens cDNA clone
 DEFINITION IMAGE:6598412 5', mRNA sequence.
 ACCESSION BUS49740
 VERSION BUS49740.1 GI:24034703
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 877)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2831 row: c column: 20
 High quality sequence stop: 659.
 Location/Qualifiers

FEATURES

source
 1. .877
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6598412"
 /tissue_type="teratocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_109"
 /notes="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
 XhoI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACAG(G). Library constructed by Ling Hong in the
 Laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

ORIGIN

Query Match 94.5%; Score 825.2; DB 5; Length 877;
 Best Local Similarity 98.1%; Pred. No. 5.3e-206;
 Matches 833; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
 QY 23 GCGACCGAGCGGTGGAGCTCCATAGCTAAAGCTTCCGAACTAAAGCAAGATCTCTT 82
 DB 1 GCGACCGAGCGGTGGAGCTCCATAGCTAAAGCTTCCGAACTAAAGCAAGATCTCTT 60
 QY 83 GCTCGTGGTTTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCATAT 142
 DB 61 GCTCGTGGTTTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAGACTCCAGGCATAT 120
 QY 143 CTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGATGATCTGGAGATGAACAGAG 202
 DB 121 CTTGAAGAACATGCTGAAGAGGAGGCAAAATGAAGAGATGATCTGGAGATGAACAGAG 180
 QY 203 GAAGAGAACCAAGCCCATTTAGCTCCCTGTCAAGAGGAAGAACCCCTGAAAAACT 262
 DB 181 GAAGAGAACCAAGCCCATTTAGCTCCCTGTCAAGAGGAAGAACCCCTGAAAAACT 240
 QY 263 GTTGTATGTGCAGCAGAGAGAAAGTGGTGAATAATTACATCTCAAAATACCACAGCTGAG 322
 DB 241 GTTGTATGTGCAGCAGAGAGAAAGTGGTGAATAATTACATCTCAAAATACCACAGCTGAG 300
 QY 323 AGAATGAGAAAGGGCTGAACGATTCATGTCCTGTGAGCTTGGAGAGTAAGAAAGCT 382
 DB 301 AGAATGAGAAAGGGCTGAACGATTCATGTCCTGTGAGCTTGGAGAGTAAGAAAGCT 360

QY 383 GCTCGGGCAGCTAGGTTTGGGATTTCTTCTCAGTTTCCAAACAAAGGCTCTGTCTCATCTGATAAC 442
 DB 361 GCTCGGGCAGCTAGGTTTGGGATTTCTTCTCAGTTTCCAAACAAAGGCTCTGTCTCATCTGATAAC 420
 QY 443 AAACCTATGTTAACTTTGGATAAGCTGAAGGAAAGAGCTCAAGAGATTTGGTTTGAATGTC 502
 DB 421 AAACCTATGTTAACTTTGGATAAGCTGAAGGAAAGAGCTCAAGAGATTTGGTTTGAATGTC 480
 QY 503 TCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAAAGAGGAGGCGCATTT 562
 DB 481 TCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGAAAAAGAGGAGGCGCATTT 540
 QY 563 GGGATTGTCTCAAGTTTCACTGGAATCTGGAATCGCAACACAGAGGATACAGAGCAAGAGAGG 622
 DB 541 GGGATTGTCTCAAGTTTCACTGGAATCGCAACACAGAGGATACAGAGCAAGAGAGG 600
 QY 623 AAAAGAGCAGAGCGCTTTGGGATTCCTGATGAAAGTTTCTGATATCTTTCTCTCCA 682
 DB 601 AAAAGAGCAGAGCGCTTTGGGATTCCTGATGAAAGTTTCTGATATCTTTCTCTCCA 660
 QY 683 GTGTTTTCATTTCTCTCTTCTTCTGTCACATATATGCCCTAAATGCACAGTCAATGTG 742
 DB 661 GTGTTTTCATTTCTCTCTTCTTCTGTCACATATATGCCCTAAATGCACAGTCAATGTG 720
 QY 743 CTTAGCTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACCTGCGGCA 802
 DB 721 CTTAGCTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACCTGCGGCA 780
 QY 803 GCAGTTTGACTATTGCTGTTTTCAGCTTTAAAGTTTGTGTTTTGTTTTGATTATGT 862
 DB 781 GCAGTTTGACTATTGCTGTTTTCAGCTTTAAAGTTGTTGTTGTTTTGATTATGT 840
 QY 863 TCGTTGTGA 871
 DB 841 GCTTGGNTA 849

RESULT 5
 BUS49740
 LOCUS AGENCOURT 7994057 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6082078
 DEFINITION 5', mRNA sequence.
 ACCESSION BUS49740.1 GI:22688271
 VERSION BUS49740
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 895)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2310 row: a column: 23
 High quality sequence stop: 650.
 Location/Qualifiers
 1. .895
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6082078"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"

FEATURES
 source
 1. .895
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6082078"
 /tissue_type="ductal carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"

[illegible]

317 ACTGAGAGATGTCAGAGAGGGCTGAACGATTCATGCTACCTGTGAGCTTGGAGAGTAAG 376
|||||
301 ACTGAGAGATGTCAGAGAGGGCTGAACGATTCATGCTACCTGTGAGCTTGGAGAGTAAG 360
|||||
377 AAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAAGTTCTTCAACAAAGGCTGTGATCT 436
|||||
361 AAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAAGTTCTTCAACAAAGGCTGTGATCT 420
|||||
437 GATAACAAACCTATGTTACTTGGATTCAGTCAAGGAAAGAGCTCAAGAGATTTGGTTTG 496
|||||
421 GATAACAAACCTATGTTACTTGGATTCAGTCAAGGAAAGAGCTCAAGAGATTTGGTTTG 480
|||||
497 AATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGGAAGAGAGAGGAG 556
|||||
481 AATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACTGGAAGAGAGAGGAG 540
|||||
557 CGATTTGGGATTTGCAAACTTCAAGTTGAGTGGAACTGGAACCAACAGAGATACAGAGCAAG 616
|||||
541 CGATTTGGGATTTGCAAACTTCAAGTTGAGTGGAACTGGAACCAACAGAGATACAGAGCAAG 600
|||||
617 AAGAGGAAAGAGACAGAGCGCTTTGGGATTCCTGATGAAAGTTCCTGATACCTTCTGT 676
|||||
601 AAGAGGAAAGAGACAGAGCGCTTTGGGATTCCTGATGAAAGTTCCTGATACCTTCTGT 660
|||||
677 TCTCCAGTGTTCCTCAATTTCTCTCTCTTCTTCTTGGTCCATATATATGCTAAATGCACAGT 736
|||||
661 TCTCCAGTGTTCCTCAATTTCTCTCTCTTCTTCTTGGGACATATATATGCTAAATGCACAGT 720
|||||
737 CATGTGCTAGCTCTGCTCGCAATGAGGAGGATGTACCCAGGTACATCCATCAACT 796
|||||
721 CAGGTGCTACCTCTGCTCGCAATGAGGAGGATGTACCCAGGTACATCCATCAACT 780
|||||
797 GCGGACGAGTTGACTTATGCTGTTTCAAGTTTAAAGTGTGTTGTTTGTGTTTGA 856
|||||
781 GCGGACGAGTTGACTTATGCTGTTTCAAGTTTAAAGTGTGTTGTTTGTGTTTGA 840
|||||
857 TTATGTTG 864
|||||
841 TTAAGGTG 848
|||||

CO648068 963 bp mRNA linear EST 23-JUL-2004
ILLUMIGEN MQO 41334 Katze_MMPB2 Macaca mulatta cDNA clone
IBIUM:25134 5' similar to Bases 527 to 886 highly similar to human
CIP29 (Hs.410597), mRNA sequence.
CO648068
CO648068.1 GI:50569562
EST.
Macaca mulatta (rhesus monkey)
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
1 (bases 1 to 963)
Katze,M.G., Thomas,M., Korth,M., Iadonato,S.P. and Magnus,C.L.
Large-scale Rhesus Macaque cDNA Sequencing
Unpublished (2003)
Contact: C. Magnus
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagnus@illumigen.com
Sequenced on 2004.06.02. 750 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCTCTACTAAGGGAACAAA
BACKWARD: CACTATAGGGCGAATGGGTA
Insert Length: 963 Std Error: 0.00

Plate: CL000345 row: H column: 08
Seq primer: CCTCTACTAAGGGAACAAA
POLYA=Yes.

FEATURES
source
1. .963
Location/Qualifiers
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBIUM:25134"
/sex="male"
/tissue_type="blood"
/cell_type="PBMC"
/lab_host="Electromax DH10B"
/clone_lib="Katze_MMPB2"
/note="Vector: pDONR 222; Site 1: Berg I; Site 2: Berg I;
Created from CloneMiner cDNA Library Construction kit
(catalog #18249-029)"

ORIGIN
Query Match 94.0%; Score 820.4; DB 7; Length 963;
Best Local Similarity 98.1%; Pred. No. 1e-204;
Matches 830; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 9 GGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTAAAGCTTGCCTGAACTAA 68
Db 3 GGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTTGAAGCTTGCCTGAACTAA 62
QY 69 AGCAAGAATGTCTTGTCTGCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACA 128
Db 63 AGCAAGAATGTCTTGTCTGCTGCTGTTGGAGACCAAGGGAATAAAGCAAGATCTTATCCACA 122
QY 129 GACTCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAAAATGAAGAGATGTACTGG 188
Db 123 GACTCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAAAATGAAGAGATGTACTGG 182
QY 189 GAGTGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCTCTGTCAGAGAGGAAGAAC 248
Db 183 GAGTGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCTCTGTCAGAGAGGAAGAAC 242
QY 249 CCCCTGAAAAAACTGTGTATGTGGCAGCAGAGAAAGTGGTGAATAATTAATCTCTGAAA 308
Db 243 CCCCTGAAAAAACTGTGTATGTGGCAGCAGAGAAAGTGGTGAATAATTAATCTCTGAAA 302
QY 309 TACCACAGACTGAGAGAAATGCAAGAGGCTGAAGGCTCAATGTACCTGTGAGCTTGG 368
Db 303 TACCACAGACTGAGAGAAATGCAAGAGGCTGAAGGCTCAATGTACCTGTGAGCTTGG 362
QY 369 AGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGTC 428
Db 363 AGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGTC 422
QY 429 TGTCTATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAAGAGCTCAAGAT 488
Db 423 TGTCTATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAAGAGCTCAAGAT 482
QY 489 TTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGA 548
Db 483 TTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACTGAAAAAGA 542
QY 549 GGAAGGAGCGATTTGGGATTTGCACAAAGTTTCAGCTGGAACCTGGAACCCACAGAGGATACAG 608
Db 543 GGAAGGAGCGATTTGGGATTTGCACAAAGTTTCAGCTGGAACCTGGAACCCACAGAGGATACAG 602
QY 609 AGGCAAGAGAGGAAAAAGAGCAGCGCTTTGGGATTCCTGATGATGAAAGTTCTCTGATA 668
Db 603 AGGCAAGAGAGGAAAAAGAGCAGCGCTTTGGGATTCCTGATGATGAAAGTTCTCTGATA 662
QY 669 CTTTCTGTTCTCCAGTGTGTTTTCATTTCTCTCTTCTTCTTCTTCTGATATATGCTTAA 728
Db 663 CTTTCTGTTCTCCAGTGTGTTTTCATTTCTCTCTTCTTCTTCTTCTGATATATGCTTAA 722

```

Qy 729 TGCAAGTCTATGTCCTTACGTCCTGCTCCCAATGAGGGAGCATGTACCCAGGTACATC 788
Db 723 TGCAAGTCTATGTCCTTACGTCCTGCTCCCAATGAGGGAGCATGTACCCAGGTACATC 782
Qy 789 CATGAAGTCTGGGAGGAGTTGACTTATTCCTGTTTTCAGCTTTAAGGTTGCTGTTT 848
Db 783 CGTGAAGTCTGGGAGGAGTTGACTTATTCCTGTTTTCAGCTTTAAGGTTGCTGTTT 842
Qy 849 GTTTTT 854
Db 843 TGTTTT 848

RESULT 8
CR610192 844 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CSOCAP004Yall of Thymus of Homo sapiens
DEFINITION CR610192
ACCESSION CR610192.1 GI:50490999
VERSION HTC; CNSLT cDNA
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 844)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue
2 (bases 1 to 844)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source Location/Qualifiers
1..844
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOCAP004Yall"
/tissue_type="Thymus"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 93.9%; Score 820; DB 3; Length 844;
Best Local Similarity 100.0%; Pred. No. 1.2e-204;
Matches 820; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 AGCTTGCCTAAAGCAAGATGCTTGTCTGCTGGTTGGAGACCAAGGGAATAAAGC 113
Db 1 AGCTTGCCTAAAGCAAGATGCTTGTCTGCTGGTTGGAGACCAAGGGAATAAAGC 60

Qy 114 AAGATCTTATCCACAGCTCCAGGCATATCTTGAGAAACATGCTGAAGAGAGGCAATG 173
Db 61 AAGATCTTATCCACAGCTCCAGGCATATCTTGAGAAACATGCTGAAGAGAGGCAATG 120

Qy 174 AAGAGATGTTACTGGGAGATGAACAGAGGAGAGAAACAAAGCCATTGCTCCCTG 233
Db 121 AAGAGATGTTACTGGGAGATGAACAGAGGAGAGAAACAAAGCCATTGCTCCCTG 180

Qy 234 TCAAGAGGAGAACCCCTGAAACAACTGTTGATGTGCGACGAGAGAAAGTGGTGA 293
Db 181 TCAAGAGGAGAACCCCTGAAACAACTGTTGATGTGCGACGAGAGAAAGTGGTGA 240

```

```

Qy 294 AAATTACATCTGAATATCACACAGCTGAGAGATGAGAGAGGGCTGACAGATTCAATG 353
Db 241 AAATTACATCTGAATATCACACAGCTGAGAGATGAGAGAGGGCTGACAGATTCAATG 300
Qy 354 TACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAG 413
Db 301 TACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTTGGGATTTCTTCAG 360
Qy 414 TTCCAAACAAAGGCTGTGTCAATGATAACAAACCTATGGTTAACTTGGATAAGCTGAAG 473
Db 361 TTCCAAACAAAGGCTGTGTCAATGATAACAAACCTATGGTTAACTTGGATAAGCTGAAG 420
Qy 474 AAAGAGCTCAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGCTGGAAGATGATG 533
Db 421 AAAGAGCTCAAGATTTGGTTTGAATGCTCTTCAATCTCCAGAAAGCTGGAAGATGATG 480
Qy 534 AGAAAGCTGAAAGAGAGAGGAGCGGATTTGGGATTTGTCACAAGTTTCAGCTGGAACTGAA 593
Db 481 AGAAAGCTGAAAGAGAGAGGAGCGGATTTGGGATTTGTCACAAGTTTCAGCTGGAACTGAA 540
Qy 594 CCACAGAGGATACAGAGGCAAGAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCCTGAT 653
Db 541 CCACAGAGGATACAGAGGCAAGAGAGGAAAGAGAGAGAGCGCTTTGGGATTTGCCTGAT 600
Qy 654 GAAAAGTTCTGTGATATCTTCTGTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCT 713
Db 601 GAAAAGTTCTGTGATATCTTCTGTCTCCAGTGTCTTCCATTTCTCTCTCTCTCTCTCTCT 660
Qy 714 ACATATATGCTTAATGACAGCTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 773
Db 661 ACATATATGCTTAATGACAGCTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Qy 774 TACCCAGGTACATCCATGAATGCGGCGACAGCTTTCAGCTTATTCGTTGTTTACGTTTAA 833
Db 721 TACCCAGGTACATCCATGAATGCGGCGACAGCTTTCAGCTTATTCGTTGTTTACGTTTAA 780
Qy 834 GGTTGTTGCTTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTAA 873
Db 781 GGTTGTTGCTTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTAA 820

RESULT 9
AF161434 910 bp mRNA linear HTC 22-MAY-2001
LOCUS Homo sapiens HSPC316 mRNA, partial cds.
DEFINITION AF161434
ACCESSION AF161434.1 GI:6841281
VERSION HTC.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 910)
AUTHORS Ye, M., Zhang, Q.H., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,
Fan, H.Y., Mao, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.
TITLE Human partial CDS from cd34+ stem cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 910)
AUTHORS Ye, M., Zhang, Q.H., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,
Fan, H.Y., Mao, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1999) Shanghai Institute of Hematology, Shanghai
Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,
Shanghai 200025, P. R. China
FEATURES
source Location/Qualifiers
1..910
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CBLAIH07"
/cell_type="cd34+ stem cell"
/tissue_type="blood"
<1..765
CDS

```

/codon_start=2 /product="HSPC316" /protein_id="AAP28994.1" /db_xref="GI:6841282" /translation="SGVRGNKMAETVELHLKLAEKOECLARGLETGKIQKDLIHR LOAYLEHAEEANEEDVLGDETESEETKPIELPVKEBPPPERTVDAAEKVKVKTIS EIPTQMRQARBFNVPVLSLEKKAARAVGISVVPNKRSVILITNLMLTWISLKE ELKDLGLNVSSIIRKSEDDKLLKRRERFIVTSSAGTGTEDTEARRRKEQSALGLP DEKFLILSLVQLCFPFELVTPMPKCTVMCLRPASQ"		843 GTTGTGTTTTTGTGTTTTTGATTATGTTGCTGTTGTAAT 878	
ORIGIN			
Query Match 93.9%; Score 820; DB 3; Length 910; Best Local Similarity 99.0%; Pred. No. 1.3e-204; Matches 867; Conservative 0; Mismatches 5; Indels 4; Gaps 4;			
QY	1	TGAGTGGAGGGTAAACAGATGCGGACCGAGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC	60
DB	4	TGGAGTGGAGGGTAAACAGATGCGGACCGAGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC	63
QY	61	CGAACTAAAGCAAGAATGCTTCTGCTGCTGTTTGGAGACCAAGGGAATAAGCAAGATCT	120
DB	64	CGAACTAAAGCAAGAATGCTTCTGCTGCTGTTTGGAGACCAAGGGAATAAGCAAGATCT	123
QY	121	TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAAAATGAAGAAG	180
DB	124	TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGGAAGAGGAGGCAAAATGAAGAAG	183
QY	181	TGTACTGGGAGATGAACACAGAGAGAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA	240
DB	184	TGTACTGGGAGATGAACACAGAGAGAGAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA	243
QY	241	GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGAGTGGTGAATAATTAC	300
DB	244	GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGAGTGGTGAATAATTAC	303
QY	301	ATCTGAAATACACAGACTGAGAGATGCGAAGAGAGGGCTGGAACGATTCATTAATGTACTGT	360
DB	304	ATCTGAAATACACAGACTGAGAGATGCGAAGAGAGGGCTGGAACGATTCATTAATGTACTGT	363
QY	361	GAGCTGGAGAGTAAGAAAGCTGCTGGGAGCAGTAGGTTGGGATTTTCAGTT-CCAA	419
DB	364	GAGCTGGAGAGTAAGAAAGCTGCTGGGAGCAGTAGGTTGGGATTTTCAGTTTCCCAA	423
QY	420	CAAAAGTCTGTCAAT-CTGATAACAAACCTATGTTAACTTGGATAGCTGAAGGAAGA	478
DB	424	CAAAAGTCTGTCAATCTGATAACAAACCTATGTTAACTTGGATAGCTTGAAGGAAGA	483
QY	479	GCTCAAGATTTT-GGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAA	537
DB	484	GCTCAAGATTTTGGTTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAA	543
QY	538	ACTGAAAAGAGGAGAGCGATTTGGGATGTGCAAGTTTCAGCTGGAACTGGAACCCAC	597
DB	544	ACTGAAAAGAGGAGAGCGATTTGGGATGTGCAAGTTTCAGCTGGAACTGGAACCCAC	603
QY	598	AGAGGATACAGAGCAAGAGAGAGAAAGAGCAGCAGCGCTTTGGGATTCCTCGATGAAA	657
DB	604	AGAGGATACAGAGGC-AAGAGAGAGAAAGAGCAGCAGCGCTTTGGGATTCCTCGATGAAA	662
QY	658	AGTTCTGTGATCTTCTGTTCTCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCACAT	717
DB	663	AGTTCTGTGATCTTCTGTTCTCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCTCTCACAT	722
QY	718	ATATGCTAAATGACAGTCAATGTGCTACGTCCTGCTCGAATGAGGGAGCATGTACC	777
DB	723	ATATGCTAAATGACAGTCAATGTGCTACGTCCTGCTCGAATGAGGGAGCATGTACC	782
QY	778	CCAGGTACATCCAGTACCTGCGCAGCAGTTGACTATTGCTGTTTCAGCTTTAAGGTT	837
DB	783	CCAGGTACATCCAGTACCTGCGCAGCAGTTGACTATTGCTGTTTCAGCTTTAAGGTT	842
QY	838	GTTGTGTTTTTGTGTTTTTGATTATGTTGCTGTTGTAAT 873	
RESULT 10			
BM460786			
LOCUS			
DEFINITION		BM460786 1070 bp mRNA linear EST 05-FEB-2002	
		AGENCOURT 6421520 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532141	
		5', mRNA sequence.	
ACCESSION			
BM460786			
VERSION		BM460786.1 GI:18509826	
KEYWORDS		EST.	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE			
AUTHORS		1 (bases 1 to 1070)	
TITLE		NIH-MGC http://mgc.nci.nih.gov/.	
JOURNAL		National Institutes of Health, Mammalian Gene Collection (MGC)	
COMMENT		Unpublished (1999)	
		Contact: Robert Strausberg, Ph.D.	
		Email: cgabbs-r@mail.nih.gov	
		Tissue Procurement: ATCC	
		cDNA Library Preparation: Life Technologies, Inc.	
		cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	
		DNA Sequencing by: Agencourt Bioscience Corporation	
		Clone distribution: MGC clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LLNL at:	
		http://image.llnl.gov	
		Plate: LLAM12214 row: o column: 22	
		High quality sequence stop: 593.	
FEATURES			
source		1..1070	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:5532141"	
		/tissue_type="leiomyosarcoma"	
		/lab_host="DH10B (phage-resistant)"	
		/clone_lib="NIH_MGC_71"	
		/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2.1 kb."	
ORIGIN			
Query Match 93.3%; Score 814.4; DB 4; Length 1070; Best Local Similarity 98.7%; Pred. No. 3.9e-203; Matches 852; Conservative 0; Mismatches 7; Indels 4; Gaps 3;			
QY	10	GGGTAAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGGCCGAATCAA	69
DB	4	GGGAAACAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTTGGCCGAATCAA	63
QY	70	GCAAGAATCTCTGCTGCTGTTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAG	129
DB	64	GCAAGAATCTCTGCTGCTGTTGGAGACCAAGGGAATAAGCAAGATCTTATCCACAG	123
QY	130	ACTCCAGGATATCTTGAAGACATGCTCAAGAGGAGGCAATGAAGAGATGATCTGGG	199
DB	124	ACTCCAGGATATCTTGAAGACATGCTCAAGAGGAGGCAATGAAGAGATGATCTGGG	183
QY	190	AGATGAACACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTCAAAGAGGAAGAAC	249
DB	184	AGATGAACACAGAGGAAGAAACAAAGCCATTGAGCTCCCTGTCAAAGAGGAAGAAC	243
QY	250	CCCTGAAAAAATCTGTTGATGTGCGACAGAGAGAAAGTGGTGAATAATACATCTGAAT	309
DB	244	CCCTGAAAAAATCTGTTGATGTGCGACAGAGAGAAAGTGGTGAATAATACATCTGAAT	303
QY	310	ACCACAGACTGAGAGATGACAGAGAGGCTGACCGATTCAATGTACCTGTGAGCTTGA	369
DB	304	ACCACAGACTGAGAGATGACAGAGAGGCTGACCGATTCAATGTACCTGTGAGCTTGA	363
QY	370	GAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCTCAGTTCCCAAAAGGTCT	429


```
RESULT 12
BX456775/c
LOCUS BX456775 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004YA11 linear EST 06-MAY-2004
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX456775
VERSION BX456775.2 GI:47071640
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1. (bases 1 to 817)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 22, 2003 this sequence version replaced gi:31034801.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 6407.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0CAP004AA06NP1&c=6407.r.

FEATURES
source
1..817
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP004YA11"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="Vector: pCMVSPORT_6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match 92.9%; Score 811.4; DB 5; Length 817;
Best Local Similarity 99.1%; Pred. No. 2.2e-202;
Matches 810; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 54 AGCTTGCAGCACTAAAGCAAGATGCTTCTGCTGCTGTTGGAGACCAAGGGAATAAGC 113
Db |||||||
QY 817 AGCTTGCAGCACTAAAGCAAGATTTCTTCTGCTGCTGTTGGAGACCAAGGGAATAAGC 758
Db |||||||
QY 114 AAGATCTTATCCAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAAAATG 173
Db |||||||
QY 757 AAGATCTTATCCAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGAGGCAAAATG 698
Db |||||||
QY 174 AAGAAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTAGCTCCCTG 233
Db |||||||
QY 697 AAGAAGATGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTAGCTCCCTG 638
Db |||||||
QY 234 TCAAGAGGAAGAACCCCTGAAAAAATCTTGTATGTTGGCAGCAGAGAGAAAGTGTGA 293
Db |||||||
QY 637 TCAAGAGGAAGAACCCCTGAAAAAATCTTGTATGTTGGCAGCAGAGAGAAAGTGTGA 578
Db |||||||
QY 294 AAATTACATCTGAATACCAAGACTCAGAGAGATGTCAGAGAGGGCTGAACGATTCAATG 353
Db |||||||
QY 577 AAATTACATCTGAATACCAAGACTCAGAGAGATGTCAGAGAGGGCTGAACGATTCAATG 518
Db |||||||
QY 354 TACTGTGAGCTTGGAGAGTAAGAAAGCTGCTGGGCAGCTAGGTTTGGGATTTCTTCAG 413
Db |||||||
QY 517 TACTGTGAGCTTGGAGAGTAAGAAAGCTGCTGGGCAGCTAGGTTTGGGATTTCTTCAG 458
Db |||||||

414 TTCCAACAAAAGGCTGTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGG 473
Db |||||||
457 TTCCAACAAAAGGCTGTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGG 398
QY 474 AAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAAGTCTGAAGATGATG 533
Db |||||||
397 AAAGAGCTCAAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAAGTCTGAAGATGATG 338
QY 534 AGAACTGAAAAGAGAGAGGCGGATTTGGGATTTGTCAAGTTTCAGCTGGAACTGGAA 593
Db |||||||
337 AGAACTGAAAAGAGAGAGGCGGATTTGGGATTTGTCAAGTTTCAGCTGGAACTGGAA 278
QY 594 CCACAGAGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 653
Db |||||||
277 CCACAGAGATACAGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 218
QY 654 GAAAAGTCTCTGATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCT 713
Db |||||||
217 GAAAAGTCTCTGATACTTTCTGTTCTCCAGTGTGTTTCCATTTCTCTCTCTCTCTCTCT 158
QY 714 ACATATATGCTTAATGACAGTATGTCCTACGCTACGCTCGCCTCGCAATGAGGAGAGATG 773
Db |||||||
157 ACATATATGCTTAATGACAGTATGTCCTACGCTACGCTCGCCTCGCAATGAGGAGAGATG 98
QY 774 TACCCAGAGTACATCCATCAACTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 833
Db |||||||
97 TACCCAGAGTACATCCATCAACTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38
QY 834 GGTGTTGTTGTTTTGTTTTGTTTTCATTATGTTGCTTGT 870
Db |||||||
37 GGTGTTGTTGTTTTGTTTTGTTTTCATTATGTTGCTTGT 1

RESULT 13
BX456776
LOCUS BX456776 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004YA11 linear EST 06-MAY-2004
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION BX456776
VERSION BX456776.2 GI:47072683
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 843)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 22, 2003 this sequence version replaced gi:31036705.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 6407.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0CAP004AA06QPI&c=6407.r.

FEATURES
source
1..843
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP004YA11"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="Vector: pCMVSPORT_6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
with a NotI-oligo(dT) primer. Five prime end enriched,
```

double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN		Query Match		Best Local Similarity		Matches		815; Conservative		0; Mismatches		4; Indels		1; Gaps		1; Length		843;		99.4%; Pred. No. 8.6e-200;		91.8%; Score 801.6; DB 5;		Length 843;	
Qy	54	AGCTTGGCGA	CTAAAGCAAGATG	CTTCTGCTGCTG	TTGGAGACCAAGGGAATAAGC	113																			
Db	1	AGCTTGGCGA	CTAAAGCAAGATG	CTTCTGCTGCTG	TTGGAGACCAAGGGAATAAGC	60																			
Qy	114	AGATCTTATCC	CAGACTCCAGCATAT	CTTGAAGAACATG	CTGAGAGAGGCAATG	173																			
Db	61	AGATCTTATCC	CAGACTCCAGCATAT	CTTGAAGAACATG	CTGAGAGAGGCAATG	120																			
Qy	174	AAGAAGATGT	ACTGGGAGATGAA	CAGAGGAGAAACAA	AGGCCCATTTGAGCTCCCTG	233																			
Db	121	AGAGATGT	ACTGGGAGATGAA	CAGAGGAGAAACAA	AGGCCCATTTGAGCTCCCTG	180																			
Qy	234	TCAAAGAGGA	AGAACCCCTGAAAA	AACTGTTGATGTGG	CAGCAGAGAGAAAGTG	293																			
Db	181	TCAAAGAGGA	AGAACCCCTGAAAA	AACTGTTGATGTGG	CAGCAGAGAGAAAGTG	240																			
Qy	294	AAATTACATCT	GAAATACACAGACT	GAGAGATGCAAGAG	AGGGCTGAACGATCAATG	353																			
Db	241	AAATTACATCT	GAAATACACAGACT	GAGAGATGCAAGAG	AGGGCTGAACGATCAATG	300																			
Qy	354	TACTGTGAGCT	TGGAGTAAGAAAG	CTGCTGGGAGCTAG	TTTGGGATTTCTTTCAG	413																			
Db	301	TACTGTGAGCT	TGGAGTAAGAAAG	CTGCTGGGAGCTAG	TTTGGGATTTCTTTCAG	360																			
Qy	414	TTCCAAACAA	AGGTCTGTCTATCT	GATAACAAACCTAT	GTGTAAGCTGAAGG	473																			
Db	361	TTCCAAACAA	AGGTCTGTCTATCT	GATAACAAACCTAT	GTGTAAGCTGAAGG	420																			
Qy	474	AAAGAGCTCA	AAAGATTTGGTTG	ATGATCTCTCAAT	CTCCAGAAAGCTGAAGATG	533																			
Db	421	AAAGAGCTCA	AAAGATTTGGTTG	ATGATCTCTCAAT	CTCCAGAAAGCTGAAGATG	480																			
Qy	534	AGAACTGAA	AAAGAGAGGCGAT	TTGGATGTGCAAG	ATTTGAGCTGGAACTGGAA	593																			
Db	481	AGAACTGAA	AAAGAGAGGCGAT	TTGGATGTGCAAG	ATTTGAGCTGGAACTGGAA	540																			
Qy	594	CCACAGAGGA	TACAGAGCAAGAA	GAGAGAGAGAGAG	CGCTTTGGGATTCCTGAT	653																			
Db	541	CCACAGAGGA	TACAGAGCAAGAA	GAGAGAGAGAGAG	CGCTTTGGGATTCCTGAT	600																			
Qy	654	GAAAAGTTCT	GATACATTTCTG	TTCTCCAGTGTTC	CAATTCCTCTTCTTGGTC	713																			
Db	601	GAAAAGTTCT	GATACATTTCTG	TTCTCCAGTGTTC	CAATTCCTCTTCTTGGTC	660																			
Qy	714	ACATATATG	CTTAATGCAAGT	CTGCTAGCTCTG	CTGCAATGAGGAGCATG	773																			
Db	661	ACATATATG	CTTAATGCAAGT	CTGCTAGCTCTG	CTGCAATGAGGAGCATG	720																			
Qy	774	TACCCAGGTA	CATCCATGAATG	CGCGCAGAGTTT	GACTTATGCTGTTTCA	833																			
Db	721	TACCCAGGTA	CATCCATGAATG	CGCGCAGAGTTT	GACTTATGCTGTTTCA	779																			
Qy	834	GGTTGTTGT	GTGTTTGTGAT	TATGTTGCTGTT	TAAT 873																				
Db	780	GGTTGTTGT	GTGTTTGTGAT	TATGTTGCTGTT	TAAT 819																				

RESULT 14
BU599301
LOCUS
DEFINITION
AGENCY 8908959 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6458323
5', mRNA sequence.
BU599301
ACCESSION
BU599301.1 GI:23251060

EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 823)
NTH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Cloning by: Agencourt Bioscience Corporation
Cloning distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2641 row: j column: 20
High quality sequence stop: 602.
Location/Qualifiers
1..823
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6458323"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_142"
/notes="Vector: pDNR-LIB; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccctctggcc). Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:
5'-AACAGCTGGTATCAACGAGGCGGCGCATTCAGCGCGG-3' and 5'-ATTCTAGAGCGGCGGCGGCGCATG-dt(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the >0.5 kb size fraction (other fractions present in NIH_MGC_141). Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."
Query Match 91.5%; Score 798.8; DB 5; Length 823;
Best Local Similarity 99.3%; Pred. No. 4.7e-199;
Matches 811; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
Qy 1 TGGAGTGGGGTAAACAAGATGGCTGGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC 60
Db 8 TGGAGTGGGGTAAACAAGATGGCTGGAGCGGTGGAGCTCCATAAGCTAAAGCTTGC 67
Qy 61 CGAACTAAAGCAAGATGTTCTGCTGCTGTTGGAGCCCAAGGAATAAGCAAGATCT 120
Db 68 CGAACTAAAGCAAGATGTTCTGCTGCTGTTGGAGCCCAAGGAATAAGCAAGATCT 127
Qy 121 TATCCACAGACTCCAGCATATCTTGAAGAACATGCTGAGAGGAGGCAATGAGAGAG 180
Db 128 TATCCACAGACTCCAGCATATCTTGAAGAACATGCTGAGAGGAGGCAATGAGAGAG 187
Qy 181 TGTACTGGGAGATGAAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAAGA 240
Db 188 TGTACTGGGAGATGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 247
Qy 241 GGAAGAACCCCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAGAGAGAGAGAGAGAG 300
Db 248 GGAAGAACCCCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAGAGAGAGAGAGAG 307
Qy 301 ATCTGAAATACACAGACTGAGAGATGCAAGAGAGGCTGAACGATTCAATGACTCTGT 360

Db	308	ATCTGAAATACCA	CAGACTGAGAGAAATG	CAGAAAGGGCTGAAC	GATTCAATG	TACCTGT	367
Qy	361	GAGCTTTGGAGAGTA	AGAAAGCTGCTCGGCAGCTAG	GTGTTGGGATTTCTT	CAGTTTCCAAC	420	
Db	368	GAGCTTTGGAGAGTA	AGAAAGCTGCTCGGCAGCTAG	GTGTTGGGATTTCTT	CAGTTTCCAAC	427	
Qy	421	AAAAGTCTGT	CAATCGATAACAAACCTAT	TGGTTAACTTTGGATAGCT	GAAGGAAAGC	480	
Db	428	AAAAGTCTGT	CAATCGATAACAAACCTAT	TGGTTAACTTTGGATAGCT	GAAGGAAAGC	487	
Qy	481	TCAAAAGATT	TGGTTTCAATATGCTCTT	CAATCTCCAGAAAGCTG	GAAGATGATG	540	
Db	488	TCAAAAGATT	TGGTTTCAATATGCTCTT	CAATCTCCAGAAAGCTG	GAAGATGATG	547	
Qy	541	GA AAAAGAGGA	AGGAGCGATT	TTGGGATTTGCA	CAAGTTTCA	GCTGGAAC	600
Db	548	GA AAAAGAGGA	AGGAGCGATT	TTGGGATTTGCA	CAAGTTTCA	GCTGGAAC	607
Qy	601	GGATACAGAGG	CAAAAGAGGAAGAGC	AGAGCGCTTTGGGAT	TGCTGATG	660	
Db	608	GGATACAGAGG	CAAAAGAGGAAGAGC	AGAGCGCTTTGGGAT	TGCTGATG	667	
Qy	661	TCCTGATAC	TTTCTGTTTCCAGTGT	TTTCCATTTTCTCT	CTCTCTTCTT	720	
Db	668	TCCTGATAC	TTTCTGTTTCCAGTGT	TTTCCATTTTCTCT	CTCTCTTCTT	727	
Qy	721	TGCTAAATG	CAAGCATGTGCTAC	GTCTGCTCGCAAT	TGAGGAGCAT	780	
Db	728	TGCTAAATG	CAAGCATGTGCTAC	GTCTGCTCGCAAT	TGAGGAGCAT	787	
Qy	781	GSTACATCCAT	-GAAC	TGCGGCAGAGTTG	ACTT	814	
Db	788	GGTACATCCAT	GNACTGCGGCAGAGTT	TGACTT		822	
RESULT 15							
BU508603							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
COMMENT							

Search completed: October 5, 2005, 00:04:23
Job time : 3489.58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:18:12 ; Search time 560.261 Seconds
(without alignments)
9224.147 Million cell updates/sec

Title: US-09-788-476A-3
Perfect score: 873
Sequence: 1 tggagtggagggtaacaaga.....tgattatgttctgttgaat 873

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_16Dec04:*
- 1: Geneseq1980s:*
 - 2: Geneseq1990s:*
 - 3: Geneseq2000s:*
 - 4: Geneseq2001as:*
 - 5: Geneseq2001bs:*
 - 6: Geneseq2002as:*
 - 7: Geneseq2002bs:*
 - 8: Geneseq2003as:*
 - 9: Geneseq2003bs:*
 - 10: Geneseq2003cs:*
 - 11: Geneseq2003ds:*
 - 12: Geneseq2004as:*
 - 13: Geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	873	100.0	873	6	ABX13936
2	873	100.0	894	6	ABX13935
3	873	100.0	1071	5	AAC88100
4	873	100.0	1154	5	AAS29109
5	873	100.0	1154	6	ABX13936
6	873	100.0	1154	10	AD25243
7	873	100.0	1520	6	ABX13935
8	853.4	97.8	891	6	ABX13935
9	733.4	84.0	3147	10	AD122609
10	690.8	79.1	2553	5	ADL62461
11	603	69.1	620	13	ACN39490
12	595.2	68.2	2553	5	ADL62461
13	498.4	57.1	1022	10	AD122423
14	465.6	53.3	471	3	AAC00738
15	451.6	51.7	558	12	ADP28791
16	336.2	38.5	463	9	ACH28217
17	334	38.3	408	5	AH81555
18	265.4	30.4	5469	4	AAK84119
19	265.4	30.4	5469	4	AAK84119
20	265.4	30.4	5469	4	AAK84119

C	21	265.4	30.4	5469	4	ABA08022	Human ova
C	22	265.4	30.4	5469	8	ABX60623	CDNA enco
C	23	265.4	30.4	5469	12	ADJ31373	Human mus
C	24	265.4	30.4	9453	4	AAK84118	Human imm
C	25	265.4	30.4	9453	4	AAK84118	Human mus
C	26	265.4	30.4	9453	4	AAK84118	Human rep
C	27	265.4	30.4	9453	4	AAK84118	Human ova
C	28	265.4	30.4	9453	8	ABX60622	CDNA enco
C	29	265.4	30.4	9453	12	ADJ31372	Human mus
C	30	258.6	29.6	266	2	AAV89379	EST clone
C	31	246.2	28.2	300	10	AD122386	Rat liver
C	32	227.2	26.0	591	2	AAK85623	Novel cDN
C	33	211.8	24.3	255	5	AAH82022	Rat diffe
C	34	196.8	22.5	232	8	AAK9101	Mouse DST
C	35	164.8	18.9	399	10	ADBS0756	Primary r
C	36	154.2	17.7	215	6	ABX69315	Novel mur
C	37	92.2	10.6	463	5	ADL41869	Human ova
C	38	55	6.4	208765	12	ADQ97430	Mouse can
C	39	55.4	6.3	1686	2	AAQ87587	DNA enco
C	40	55	6.3	399	5	AAK85623	DNA enco
C	41	55	6.3	399	5	AAK85623	DNA enco
C	42	54.4	6.2	1997	13	ADR07778	Full leng
C	43	54	6.2	5659	6	ABL32375	Human imm
C	44	54	6.2	5659	6	ABL34487	Human met
C	45	54	6.2	5659	7	ADS99748	Complemen

ALIGNMENTS

RESULT 1
ABX13936
ID ABX13936 standard; DNA; 873 BP.

XX AC ABX13936;
XX DT 03-MAR-2003 (first entry)
XX DE DNA encoding novel human protein HCC-1, PCR extended form.

XX KW Human; hcc-1; human hepatocellular carcinoma tissue; HCC-M cell line;

XX KW cancer; hepatocellular carcinoma; antisense gene therapy; gene; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 20..652
XX FT /*tag= a
XX FT /product= "HCC-1"

US2002107190-A1.
08-AUG-2002.
21-FEB-2001; 2001US-00788476.
25-FEB-2000; 2000US-0185116P.

60/135,116

(CHUN/) CHUNG C M.
(CHAN/) CHAN L.
(OUKK/) OU K.
(ONGS/) ONG S.
(SEOW/) SEOW T K.
(LIAN/) LIANG C R.
(CHOO/) CHOONG M L.
(TANL/) TAN L K.

Chung CM, Chan L, Ou K, Ong S, Seow TK, Liang CR, Choong ML;
Tan LK;

WPI; 2002-697878/75.

New nucleic acid which is differentially expressed in human

QY 121 TATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGGAGGCAAAATGAAGA 180
Db |||||||
QY 285 TATCCACAGACTCCAGGCATATCTTGAAGAAACATGCTGAAGAGGAGGCAAAATGAAGA 344
Db |||||||
QY 181 TGTAATGGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCTCAAGA 240
Db |||||||
QY 345 TGTAATGGGAGATGAACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCTCAAGA 404
QY 241 GGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAAGATTAC 300
Db |||||||
QY 405 GGAAGAACCCCTGAAAGAACTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAAGATTAC 464
QY 301 ATCTGAAATACACAGACTGAGAGATGCGAAGAGAGGCTGGAACGATTCAATCTACCTGT 360
Db |||||||
QY 465 ATCTGAAATACACAGACTGAGAGATGCGAAGAGAGGCTGGAACGATTCAATCTACCTGT 524
QY 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAAC 420
Db |||||||
QY 525 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCAGCTAGGTTTGGGATTTCTTCAGTTTCCAAC 584
QY 421 AAAAGTCTGATCTGATTAACAAACCTATGGTTAACTTGGATAAGCTGAAGAAAGAGC 480
Db |||||||
QY 585 AAAAGTCTGATCTGATTAACAAACCTATGGTTAACTTGGATAAGCTGAAGAAAGAGC 644
QY 481 TCAAGATTTGGTTGAACTGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACT 540
Db |||||||
QY 645 TCAAGATTTGGTTGAACTGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAAACT 704
QY 541 GAAAGAGGAGGAGGAGGATTTGGGATTTGTCAAGTTTCAGCTGGAACTGGAAACACAGA 600
Db |||||||
QY 705 GAAAGAGGAGGAGGAGGATTTGGGATTTGTCAAGTTTCAGCTGGAACTGGAAACACAGA 764
QY 601 GATACAGAGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
Db |||||||
QY 765 GATACAGAGGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 824
QY 661 TCCTGATACCTTCTGTTCTCCAGTGTTCCTCAATTTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db |||||||
QY 825 TCCTGATACCTTCTGTTCTCCAGTGTTCCTCAATTTCTCTCTCTCTCTCTCTCTCTCTCT 884
QY 721 TGCTAAATGCACAGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db |||||||
QY 885 TGCTAAATGCACAGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
QY 781 GGTACATCCATGAACCTGCGCAGAGTTGACATTTATGCTGTTTACGTTTAAAGTTGTT 840
Db |||||||
QY 945 GGTACATCCATGAACCTGCGCAGAGTTGACATTTATGCTGTTTACGTTTAAAGTTGTT 1004
QY 841 GTGTTTTGTTTTGTTGATTATGTTGTTGTTAAT 873
Db |||||||
QY 1005 GTGTTTTGTTTTGTTGATTATGTTGTTGTTAAT 1037

RESULT 4
ID AAS29109 standard; cdna; 1154 BP.
XX AC AAS29109;
XX AC AAS29109;
DT 21-NOV-2001 (first entry)
DE cdna encoding for human DNA-binding protein #80.
XX Human; DNA-binding protein; histone; chromo domain protein;
KW chromatin organisation modifier; Y-box binding protein; DNA organisation;
KW gene transcription; malignant disease; autoimmune disorder;
KW rheumatic disease; genetic abnormality; infectious disease;
KW neurological disorder; gene therapy; immunomodulatory; anti-HIV;
KW anti rheumatic; anti microbial; cytostatic; ss.
XX Homo sapiens.
XX WO200155162-A1.

XX 02-AUG-2001.
PD 17-JAN-2001; 2001WO-US001305.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184654P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.

not entered
60 lines, 005
Not entered
checked here

29-SEP-2000; 2000US-0236367P.
29-SEP-2000; 2000US-0236368P.
29-SEP-2000; 2000US-0236369P.
29-SEP-2000; 2000US-0236370P.
02-OCT-2000; 2000US-0236802P.
02-OCT-2000; 2000US-0237037P.
02-OCT-2000; 2000US-0237038P.
02-OCT-2000; 2000US-0237039P.
02-OCT-2000; 2000US-0237040P.
13-OCT-2000; 2000US-0239935P.
13-OCT-2000; 2000US-0239937P.
20-OCT-2000; 2000US-0240960P.
20-OCT-2000; 2000US-0241221P.
20-OCT-2000; 2000US-0241785P.
20-OCT-2000; 2000US-0241786P.
20-OCT-2000; 2000US-0241787P.
20-OCT-2000; 2000US-0241809P.
20-OCT-2000; 2000US-0241826P.
01-NOV-2000; 2000US-0244617P.
08-NOV-2000; 2000US-0246474P.
08-NOV-2000; 2000US-0246475P.
08-NOV-2000; 2000US-0246476P.
08-NOV-2000; 2000US-0246477P.
08-NOV-2000; 2000US-0246478P.
08-NOV-2000; 2000US-0246523P.
08-NOV-2000; 2000US-0246524P.
08-NOV-2000; 2000US-0246525P.
08-NOV-2000; 2000US-0246526P.
08-NOV-2000; 2000US-0246527P.
08-NOV-2000; 2000US-0246528P.
08-NOV-2000; 2000US-0246532P.
08-NOV-2000; 2000US-0246609P.
08-NOV-2000; 2000US-0246610P.
08-NOV-2000; 2000US-0246611P.
08-NOV-2000; 2000US-0246613P.
17-NOV-2000; 2000US-0249207P.
17-NOV-2000; 2000US-0249208P.
17-NOV-2000; 2000US-0249209P.
17-NOV-2000; 2000US-0249210P.
17-NOV-2000; 2000US-0249211P.
17-NOV-2000; 2000US-0249212P.
17-NOV-2000; 2000US-0249213P.
17-NOV-2000; 2000US-0249214P.
17-NOV-2000; 2000US-0249215P.
17-NOV-2000; 2000US-0249216P.
17-NOV-2000; 2000US-0249217P.
17-NOV-2000; 2000US-0249218P.
17-NOV-2000; 2000US-0249244P.
17-NOV-2000; 2000US-0249245P.
17-NOV-2000; 2000US-0249264P.
17-NOV-2000; 2000US-0249265P.
17-NOV-2000; 2000US-0249297P.
17-NOV-2000; 2000US-0249299P.
01-DEC-2000; 2000US-0249300P.
01-DEC-2000; 2000US-0250160P.
05-DEC-2000; 2000US-0250391P.
05-DEC-2000; 2000US-0251030P.
05-DEC-2000; 2000US-0251988P.
05-DEC-2000; 2000US-0256719P.
06-DEC-2000; 2000US-0251479P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
08-DEC-2000; 2000US-0251990P.
11-DEC-2000; 2000US-0254097P.
05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
XX
XX
XX

DR WPI; 2001-465557/50.
XX P-PSDB; AAU18233.
XX Nucleic acid molecules encoding human secreted chromosomal binding
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers.
XX Claim 4; SEQ ID NO 90; 561pp; English.
XX The present invention relates to the isolation of novel DNA-binding
CC proteins (AAU18154-AAU18281), and cDNA and genomic sequences encoding for
CC these proteins. DNA-binding proteins such as histones, chromo (chromatin
CC organisation modifier) domain proteins, and Y-box binding proteins may
CC contribute to diseases resulting from aberrant DNA organisation and/or
CC gene transcription. The sequences of the invention are useful in
CC screening assays to identify antagonists and/or agonists that may enhance
CC or block activities mediated by DNA-binding proteins. Blockers of DNA-
CC binding proteins may be useful in treating disorders such as malignant
CC diseases (e.g. cancer), autoimmune disorders (e.g. diabetes mellitus),
CC rheumatic diseases (e.g. rheumatoid arthritis), genetic abnormalities
CC (e.g. cystic fibrosis), infectious diseases (e.g. HIV) and neurological
CC disorders (e.g. Alzheimer's disease). The polynucleotide sequences of the
CC invention may also be used in gene therapy. AAS29030-AAS29157 represent
CC cDNA sequences encoding for novel DNA-binding proteins. Note: The
CC cDNA sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1154 BP; 394 A; 186 C; 291 G; 281 T; 0 U; 2 Other;
Query Match 100.0%; Score 873; DB 5; Length 1154;
Best Local Similarity 100.0%; Pred. No. 4.5e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGAGTGGGGTAAACAAGATGCGACCGAGAGCGTGGAGCTCCATAAGCTAAAGCTTGC 60
DB 182 TGGAGTGGGGGTAAACAAGATGCGACCGAGAGCGTGGAGCTCCATAAGCTAAAGCTTGC 241
QY 61 CGAACTAAAGCAAGAATGCTTCTCGTGGTGGAGACCAAGGGAATAAGCAAGATCT 120
DB 242 CGAACTAAAGCAAGAATGCTTCTCGTGGTGGAGACCAAGGGAATAAGCAAGATCT 301
QY 121 TATCCACAGACTCAGGCATATCTTGAAGAACATCTGAAGAGAGGCAATGAGAGGA 180
DB 302 TATCCACAGACTCAGGCATATCTTGAAGAACATCTGAAGAGAGGCAATGAGAGGA 361
QY 181 TGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAGA 240
DB 362 TGTACTGGGAGATGAACACAGAGGAAGAAACAAAGCCCATTTGAGCTCCCTGTCAAGA 421
QY 241 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGTGTGAAATTTAC 300
DB 422 GGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAAGTGTGAAATTTAC 481
QY 301 ATCTGAAATACCAAGACTCAGAGATGAGAGAGGCTGAACGATTAATGACCTGT 360
DB 482 ATCTGAAATACCAAGACTCAGAGATGAGAGAGGCTGAACGATTAATGACCTGT 541
QY 361 GAGCTTGGAGATGAAGAAAGCTGCTCGGCAGCTAGTGTGGGATTTCTTCAGTTCCAAC 420
DB 542 GAGCTTGGAGATGAAGAAAGCTGCTCGGCAGCTAGTGTGGGATTTCTTCAGTTCCAAC 601
QY 421 AAAAGGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAAGAGC 480
DB 602 AAAAGGTCTGTCATCTGATAACAAACCTATGTTAACTTGGATAAGCTGAAGGAAAGAGC 661
QY 481 TCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACT 540
DB 662 TCAAGATTTGGTTGAATGCTCTTCAATCTCCAGAAAGTCTGAAGATGATGAGAACT 721
QY 541 GAAAAGAGGAGGAGCGGATTTGGATTTGTCAAGTTTCAAGTTGAGTGGAACTGAACACAGA 600
DB 722 GAAAAGAGGAGGAGCGGATTTGGATTTGTCAAGTTTCAAGTTGAGTGGAACTGAACACAGA 781

PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249445P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.

PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764845.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Barash SC;
PI WPI; 2003-605749/57.
DR P-FSDB; ADC25371.
XX

New DNA-binding proteins and gene encoding them, useful for diagnosing, PT treating and/or preventing e.g. neurological, inflammatory, infectious, PT cardiovascular, autoimmune, respiratory, neoplastic or digestive PT diseases.
XX

PS Claim 1; SEQ ID NO 90; 226pp; English.

XX The invention relates to an isolated nucleic acid molecule (cDNA) CC encoding a human extracellular matrix protein, representing one of 161 CC novel genes. Also included are recombinant vectors, host cells CC (expressing the protein), the extracellular matrix proteins (including CC their fragments, epitopes and homologues), an isolated antibody that CC binds specifically to the protein, diagnosing a pathological condition or CC susceptibility to a pathological condition (comprising determining the CC presence or absence of a mutation in the nucleic acid and diagnosing a CC condition based on the presence or absence of the mutation), diagnosing a CC pathological condition or susceptibility to a pathological condition CC (comprising determining the presence or amount of expression of the CC protein in a biological sample and diagnosing a condition based on the CC presence or amount of expression of the protein), preventing, treating or CC ameliorating a medical condition by administering the nucleic acid or CC protein to a mammalian subject, identifying a binding partner to the CC protein, the gene corresponding to the cDNA sequence, and identifying an CC activity in a biological assay (comprising expressing the nucleic acid in CC a cell, isolating the supernatant, detecting an activity in a biological CC assay and identifying the protein in the supernatant having the CC activity). The nucleic acids and proteins display the following CC activities Cytostatic, antibacterial, Virucide, Neuroprotective,

Query Match 100.0%; Score 873; DB 10; Length 1154;
Best Local Similarity 100.0%; Pred. No. 4.5e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGTGGGGTAAACAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTC 60
DB 182 TGGAGTGGGGTAAACAGATGGCGACCGAGCGGTGGAGCTCCATAAGCTTAAGCTTC 241
QY 61 CGAACTAAAGCAAGATGCTTGTCTGCTGTTGGAGACCAAGGGGAATAAGCAAGATCT 120
DB 242 CGAACTAAAGCAAGATGCTTGTCTGCTGTTGGAGACCAAGGGGAATAAGCAAGATCT 301
QY 121 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGCGCAATGAAGA 180
DB 302 TATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGCGCAATGAAGA 361
QY 181 TGTACTGGGAGATGAACAG 240
DB 362 TGTACTGGGAGATGAACAG 421
QY 241 GGAAGAACCCCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAGAGAGAGAGAGAGAG 300
DB 422 GGAAGAACCCCTGAAAAAATCTTGTATGTGGCAGCAGAGAGAGAGAGAGAGAGAGAG 481
QY 301 ATCTGAAATACCAAGACTGAGAGAGATGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 482 ATCTGAAATACCAAGACTGAGAGAGATGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541

Qy 361 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 420
Db 542 GAGCTTGGAGAGTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAAC 601
Qy 421 AAAAGGCTCTCATCTGATACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAGAGC 480
Db 602 AAAAGGCTCTCATCTGATACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAGAGC 661
Qy 481 TCAAAGATTTGGTTGAATGTCTTCAATCTCCAGAAAGCTGAAGATGATGAGAACT 540
Db 662 TCAAAGATTTGGTTGAATGTCTTCAATCTCCAGAAAGCTGAAGATGATGAGAACT 721
Qy 541 GAAAAGAGGAAGAGCGATTTGGATTGTCAAGTTTCAGCTGGAACCTGGAACCAAGA 600
Db 722 GAAAAGAGGAAGAGCGATTTGGATTGTCAAGTTTCAGCTGGAACCTGGAACCAAGA 781
Qy 601 GGATACAGAGCAAGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 782 GGATACAGAGCAAGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
Qy 661 TCCTGATACATTTCTGTTCTCCAGTGTCTTCAATTTCTCTCTCTCTCTCTCTCTCTCT 720
Db 842 TCCTGATACATTTCTGTTCTCCAGTGTCTTCAATTTCTCTCTCTCTCTCTCTCTCTCT 901
Qy 721 TGCTTAAATGCACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGAGATGTACCCCA 780
Db 902 TGCTTAAATGCACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGAGATGTACCCCA 961
Qy 781 GGTACATCCATGAATCTCGGAGAGAGTTGACTTATGCTGTTTACGCTTTAAGTTGTT 840
Db 962 GGTACATCCATGAATCTCGGAGAGAGTTGACTTATGCTGTTTACGCTTTAAGTTGTT 1021
Qy 841 GTCTTTTGTGTTTGTATGATGTTGCTGTTAAT 873
Db 1022 GTGTTTTGTTTTGATTTGCTGTTGTTAAT 1054

RESULT 7

ABN59623
ID ABN59623 standard; cDNA; 1520 BP.
AC AC
ABN59623;
DT 28-JUN-2002 (first entry)
XX Novel human coding sequence SEQ ID NO: 34.
XX Human; antianaemic; vulnery; antiinflammatory; immunomodulator;
KW antiinfertility; cerebroprotective; cytotatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag; gene; ss.
XX Homo sapiens.
OS OS
XX WO200222660-A2.
DN 21-MAR-2002.
XX 10-SEP-2001; 2001WO-US026015.
XX 11-SEP-2000; 2000US-00659671.
PR (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX WPI; 2002-292408/33.
DR P-PSDB; ABB97210.
XX An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX

PS CC The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a coding sequence of the
CC invention
XX
SQ Sequence 1520 BP; 442 A; 310 C; 423 G; 345 T; 0 U; 0 Other;
Query Match 100.0%; Score 873; DB 6; Length 1520;
Best Local Similarity 100.0%; Pred. No. 5.1e-232;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGGAGTGAAGGGTAAACAAGATGCGACCGAGAGCGTGGAGCTCCATAAGCTAAAGCTTGC 60
Db 617 TGGAGTGAAGGGTAAACAAGATGCGACCGAGAGCGTGGAGCTCCATAAGCTAAAGCTTGC 676
Qy 61 CGAACTAAAGCAAGAATGTCTTCTGCTGTTTGGAGACCAAGGAATAAGCAAGATCT 120
Db 677 CGAACTAAAGCAAGAATGTCTTCTGCTGTTTGGAGACCAAGGAATAAGCAAGATCT 736
Qy 121 TATCCACAGACTCCAGGCATATCTTGAAGAACAATGCTGAAGAGGAGGCAATGAAGAAGA 180
Db 737 TATCCACAGACTCCAGGCATATCTTGAAGAACAATGCTGAAGAGGAGGCAATGAAGAAGA 796
Qy 181 TGTACTGGGAGATGAAACAGAGGAAGAAGAAACAAGGCCCATTTGAGTCTCTGTTCAAGA 240
Db 797 TGTACTGGGAGATGAAACAGAGGAAGAAGAAACAAGGCCCATTTGAGTCTCTGTTCAAGA 856
Qy 241 GGAAGAACCCCTGAAAGAACTGTGATGTGGCAGCAGAGAGAAAGTGGTGAAGATTTAC 300
Db 857 GGAAGAACCCCTGAAAGAACTGTGATGTGGCAGCAGAGAGAAAGTGGTGAAGATTTAC 916
Qy 301 ATCTGAAATACCAAGACTGAGAGAAATGCAAGAGAGGCTGAACGATTTCAATGTACCTGT 360
Db 917 ATCTGAAATACCAAGACTGAGAGAAATGCAAGAGAGGCTGAACGATTTCAATGTACCTGT 976
Qy 361 GAGCTTGGAGATTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAC 420
Db 977 GAGCTTGGAGATTAAGAAAGCTGCTCGGCGAGCTAGGTTGGGATTTCTTCAGTTCCAAC 1036
Qy 421 AAAAGGCTCTCATCTGATACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAGAGC 480
Db 1037 AAAAGGCTCTCATCTGATACAAACCTATGGTTAACTTGGATAAGCTGAAGGAAGAGC 1096
Qy 481 TCAAAGATTTGGTTGAATGTCTTCAATCTCCAGAAAGCTCTGAAGATGATGAGAACT 540
Db 1097 TCAAAGATTTGGTTGAATGTCTTCAATCTCCAGAAAGCTCTGAAGATGATGAGAACT 1156
Qy 541 GAAAAGAGGAAGAGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAACCAAGA 600
Db 1157 GAAAAGAGGAAGAGCGATTTGGGATTTGTCAAGTTTCAGCTGGAACCTGGAACCAAGA 1216
Qy 601 GGATACAGAGCAAGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 1217 GGATACAGAGCAAGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1276
Qy 661 TCCTGATACATTTCTGTTCTCCAGTGTCTTCAATTTCTCTCTCTCTCTCTCTCTCTCT 720
Db 1277 TCCTGATACATTTCTGTTCTCCAGTGTCTTCAATTTCTCTCTCTCTCTCTCTCTCTCT 1336
Qy 721 TGCTTAAATGCACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGAGATGTACCCCA 780
Db 1337 TGCTTAAATGCACAGTCATGTGCTACGTCTGCTCGCAATGAGGAGAGATGTACCCCA 1396
Qy 781 GGTACATCCATGAATCTCGGAGAGAGTTTGAATTTGCTGTTTACGCTTTAAGTTGTT 840

CC marker corresponds to a secreted protein or to a transcribed
CC polynucleotide or its portion. The level of expression of the marker is
CC assessed by detecting the presence in the sample, a protein or protein
CC fragment corresponding to the marker. The presence of protein or protein
CC fragment is detected using an antibody that specifically binds with the
CC protein or protein fragment. Alternatively, the level of expression of
CC the marker is assessed by detecting the presence of a transcribed
CC polynucleotide which anneals with the marker or anneals with a portion of
CC the polynucleotide comprising the marker, under stringent conditions. The
CC marker is also used for monitoring the progression of ovarian cancer in a
CC patient which involves detecting expression of the marker in a patient
CC sample at a first point in time, repeating the method at a subsequent
CC time and comparing the level of expression. The method is carried out
CC using an ovarian tissue sample. A composition comprising a marker,
CC polypeptide or antibody of the invention is used to treat ovarian cancer.
CC This sequence represents a human ovarian cancer DNA marker of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2553 BP; 690 A; 597 C; 570 G; 696 T; 0 U; 0 Other;

Query Match 79.1%; Score 690.8; DB 5; Length 2553;
Best Local Similarity 99.7%; Pred. No. 3e-181;
Matches 692; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 93 TGGAGACCAAGGGAATTAAGCAAGATCTTATCCACAGACTCCAGGCATCTTTGAAGAAC 152
DB |||||||
QY 715 TGGAGACCAAGGGAATTAAGCAAGATCTTATCCACAGACTCCAGGCATCTTTGAAGAAC 774
DB |||||||
QY 153 ATGCTGAAGAGGAGGAAATGAAGAGATGTACTGGGAGATGAAACAGAGGAAGAGAAA 212
DB |||||||
QY 775 ATGCTGAAGAGGAGGAAATGAAGAGATGTACTGGGAGATGAAACAGAGGAAGAGAAA 834
DB |||||||
QY 213 CAAAGCCCATTTGAGTCCCTGTCTCAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGG 272
DB |||||||
QY 835 CAAAGCCCATTTGAGTCCCTGTCTCAAGAGGAAGAACCCCTGAAAGAACTGTTGATGTGG 894
DB |||||||
QY 273 CAGCAGAGAGAAAGTGGTGAAATTTACATCTGAAATACACAGACTGAGAGAAATGCGA 332
DB |||||||
QY 895 CAGCAGAGAGAAAGTGGTGAAATTTACATCTGAAATACACAGACTGAGAGAAATGCGA 954
DB |||||||
QY 333 AGAGGGCTGAACGATTCATGTTACCTGTGAGCTGGAGAGTAAGAAGCTGCTCGGCGAG 392
DB |||||||
QY 955 AGAGGGCTGAACGATTCATGTTACCTGTGAGCTGGAGAGTAAGAAGCTGCTCGGCGAG 1014
DB |||||||
QY 393 CTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAGGTCTGTCTCATCTGATAACAAACCTATGG 452
DB |||||||
QY 1015 CTAGGTTTGGGATTTCTTCAGTTTCCAAACAAAGGTCTGTCTCATCTGATAACAAACCTATGG 1074
DB |||||||
QY 453 TTAACCTTGGATTAAGCTGAAGAAAGAGCTCAAGATTTGGTTTGAATGTCTCTTCAATCT 512
DB |||||||
QY 1075 TTAACCTTGGATTAAGCTGAAGAAAGAGCTCAAGATTTGGTTTGAATGTCTCTTCAATCT 1134
DB |||||||
QY 513 CCAGAAAGCTGGAAGATGATGAGAACTGAAAGAGGAAGGAGCGATTTGGGATTTGCA 572
DB |||||||
QY 1135 CCAGAAAGCTGGAAGATGATGAGAACTGAAAGAGGAAGGAGCGATTTGGGATTTGCA 1194
DB |||||||
QY 573 CAAGTTCCAGCTGGAACCTGGAACCAACAGAGGATACAGAGGCAAGAGAGAAAGAGCAG 632
DB |||||||
QY 1195 CAAGTTCCAGCTGGAACCTGGAACCAACAGAGGATACAGAGGCAAGAGAGAAAGAGCAG 1254
DB |||||||
QY 633 AGGCTTTGGGATTTGCTGATGAAAGTTCTGTGATCTTCTGTTCTCCAGTGTCTTCCA 692
DB |||||||
QY 1255 AGGCTTTGGGATTTGCTGATGAAAGTTCTGTGATCTTCTGTTCTCCAGTGTCTTCCA 1314
DB |||||||
QY 693 TTTCTCTCTCTTCTTCTTGGTTCACATATATGCTTAATGACAGTTCATGTCCTACGTCCT 752
DB |||||||
QY 1315 TTTCTCTCTCTTCTTCTTGGTTCACATATATGCTTAATGACAGTTCATGTCCTACGTCCT 1374
DB |||||||
QY 753 GCCTTCGAATGAGGAGCATGTATCCCCAGGTACA 786
DB |||||||
QY 1375 GCCTTCGAATGAGGAGCATGTATCCCCAGGTACA 1408
DB |||||||

RESULT 11

ACN39490
ID ACN39490 standard; cDNA; 620 BP.

XX
AC ACN39490;

XX
DT 18-NOV-2004 (first entry)

XX
DE Tumour-associated antigenic target (TAT) cDNA DNA325701, SEQ ID NO:3662.

XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic; gene; ss.

XX
OS Homo sapiens.

XX
PN WO2004030615-A2.

XX
PD 15-APR-2004.

XX
PF 29-SEP-2003; 2003WO-US028547.

XX
PR 02-OCT-2002; 2002US-0414971P.

XX
PA (GETH) GENENTECH INC.

XX
PI Wu TD, Zhang Z, Zhou Y;

XX
DR WPI; 2004-347921/32.

XX
PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.

XX
PS Claim 1; SEQ ID NO 3662; 7273pp; English.

XX
CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention

XX
SQ Sequence 620 BP; 189 A; 107 C; 154 G; 170 T; 0 U; 0 Other;

Query Match 69.1%; Score 603; DB 13; Length 620;
Best Local Similarity 100.0%; Pred. No. 4.2e-157;
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GGCAGCAGAGAGAAAGTGGTGAATAATTACATCTGAAATACCAACAGACTGAGAGATGCA 330
DB |||||||

DB 1 GGCAGCAGAGAGAAAGTGGTGAATAATTACATCTGAAATACCAACAGACTGAGAGATGCA 60

```
QY 331 GAAGAGGCTGAACGATTCAATGTACCTGTGAGCTTGGAGTAAGAAAGCTGCTCGGC 390
DB 61 GAAGAGGCTGAACGATTCAATGTACCTGTGAGCTTGGAGTAAGAAAGCTGCTCGGC 120
QY 391 AGCTAGGTTGGGATTTCTTCAGTTCGAAACAAAGGCTGTGATCAACAACCTAT 450
DB 121 AGCTAGGTTGGGATTTCTTCAGTTCGAAACAAAGGCTGTGATCAACAACCTAT 180
QY 451 GGTAACTTGATAGCTGAAGGAAGAGCTCAAGATTTGGTTTGAATGCTCTCAAT 510
DB 181 GGTAACTTGATAGCTGAAGGAAGAGCTCAAGATTTGGTTTGAATGCTCTCAAT 240
QY 511 CTCAGAAAGTCTGAAGATGATGAAACTCAAAAGAGGAGGCGATTGGGATTGT 570
DB 241 CTCAGAAAGTCTGAAGATGATGAAACTCAAAAGAGGAGGCGATTGGGATTGT 300
QY 571 CACAAGTTCACTGGAATCGAAACACAGAGGATACAGAGGCAAGAGGAAAGAGC 630
DB 301 CACAAGTTCACTGGAATCGAAACACAGAGGATACAGAGGCAAGAGGAAAGAGC 360
QY 631 AGAGGCTTTGGGATTCGCTGATGAAGATTCTCTGATATTTCTGTTCTCCAGTGTTC 690
DB 361 AGAGGCTTTGGGATTCGCTGATGAAGATTCTCTGATATTTCTGTTCTCCAGTGTTC 420
QY 691 CATTTCTCTCTCTCTCTCTCTCTCAATATATGCTTAAATGCACAGTCAATGCTACGTC 750
DB 421 CATTTCTCTCTCTCTCTCTCTCTCAATATATGCTTAAATGCACAGTCAATGCTACGTC 480
QY 751 CTGCTCGCAATGAGGAGCATGTACCCAGGTATCATCCATGAATCGGCGAGATTG 810
DB 481 CTGCTCGCAATGAGGAGCATGTACCCAGGTATCATCCATGAATCGGCGAGATTG 540
QY 811 ACTTATGCTGTTTCACTTAAAGTCTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGT 870
DB 541 ACTTATGCTGTTTCACTTAAAGTCTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGT 600
QY 871 AAT 873
DB 601 AAT 603

RESULT 12
ADL62461/C
ID ADL62461 standard; DNA; 2553 BP.
XX AC ADL62461;
XX DT 20-MAY-2004 (first entry)
XX DE Human ovarian cancer DNA marker #20673.
XX KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX OS Homo sapiens.
XX PN WO200170979-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-0009126.
XX PR 21-MAR-2000; 2000US-0191031P.
XX PR 25-MAY-2000; 2000US-0207124P.
XX PR 15-JUN-2000; 2000US-0211940P.
XX PR 07-JUL-2000; 2000US-0216820P.
XX PR 25-JUL-2000; 2000US-0220661P.
XX PR 21-DEC-2000; 2000US-0257672P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lee J, Lillie J;
```

DR WPI; 2001-611502/70.

XX Novel isolated nucleic acid molecules (markers) overexpressed in ovarian cancer cells as compared to their normal non-cancerous ovarian cells are used to characterize stage, grade, histological type of ovarian cancer.

XX Disclosure; SEQ ID NO 20673; 106pp; English.

XX The invention relates to nucleic acid markers which are overexpressed in ovarian cancer cells as compared to their expression in normal (i.e. non-cancerous) ovarian cells. The invention also relates to polypeptides encoded by the markers, antibodies that selectively bind to the polypeptides, a method of inhibiting ovarian cancer in a patient at risk of developing ovarian cancer involving inhibiting expression of a gene corresponding to a marker of the invention and a method of treating a patient afflicted with ovarian cancer comprising providing to cells of the patient an antisense oligonucleotide complementary to a marker of the invention. The markers are useful for assessing if a patient is afflicted with ovarian cancer, which involves comparing the level of expression of the marker in a patient sample and a normal level of expression of the marker in a control non-ovarian cancer sample. A difference between the expression levels indicates ovarian cancer. The level of expression of a marker corresponds to a secreted protein or to a transcribed polynucleotide or its portion. The level of expression of the marker is assessed by detecting the presence in the sample, a protein or protein fragment corresponding to the marker. The presence of protein or protein fragment is detected using an antibody that specifically binds with the protein or protein fragment. Alternatively, the level of expression of the marker is assessed by detecting the presence of a transcribed polynucleotide which anneals with the marker or anneals with a portion of the polynucleotide comprising the marker, under stringent conditions. The marker is also used for monitoring the progression of ovarian cancer in a patient which involves detecting expression of the marker in a patient sample at a first point in time, repeating the method at a subsequent time and comparing the level of expression. The method is carried out using an ovarian tissue sample. A composition comprising a marker, polypeptide or antibody of the invention is used to treat ovarian cancer. This sequence represents a human ovarian cancer DNA marker of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 2553 BP; 690 A; 597 C; 570 G; 696 T; 0 U; 0 Other;

Query Match 68.2%; Score 595.2; DB 5; Length 2553;
Best Local Similarity 96.9%; Pred. No. 1.2e-154;
Matches 628; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

QY 202 GGAAGAAAGAAACAAAGCCATTTGAGCTCCCTGTCAAAGAGGAGAAACCCCTG-AAAAA 260

DB 695 GGAAGAAAGAAACAAAGCCATTTGAGCTCCCTGTCAAAGAGGAGAAACCCCTGAAAAA 636

QY 261 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATACATCTGAATACACAGACTG 320

DB 635 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATACATCTGAATACACAGACTG 576

QY 321 AGAAGATGCAGAGAGGGCTGAACGATTCATGTACCTGTGAGCTTGGAGTAGAAG 380

DB 575 AGAAGATGCAGAGAGGGCTGAACGATTCATGTACCTGTGAGCTTGGAGTAGAAG 516

QY 381 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTTCAAACAAAGGCTGTCTCATCTGATA 440

DB 515 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTTCAAACAAAGGCTGTCTCATCTGATA 456

QY 441 ACAAACCTATGTTTAACTTGGATAAGCTGAAGGAAAGAGCTCAAGATTTTGGTTGAATG 500

DB 455 ACAAACCTATGTTTAACTTGGATAAGCTGAAGGAAAGAGCTCAAGATTTTGGTTGAATG 396

QY 501 TCTCTCAATCTCCAGAAAGTCTGAAGATGATGAATACTGAAAAAGAGGAGGCGAT 560

DB 395 TCTCTCAATCTCCAGAAAGTCTGAAGATGATGAATACTGAAAAAGAGGAGGCGAT 336

QY 561 TTGGATTGTCAAGTTTCAAGTTCAGCTGGAATGAAACACAGAGGATACAGAGCAAGAGA 620

Db 335 TTGGATTGTCACAAAGTTACAGTGGAACTGGAACCCAGAGGATACAGAGGCAAGAGA 276
Qy 621 GGAAGAGAGAGAGCGCTTTGGGATTCCTGATGAAAGTTCCTGATACCTTCTGTTCTC 680
Db 275 GGAAGAGAGAGAGCGCTTTGGGATTCCTGATGAAAGTTCCTGATACCTTCTGTTCTC 216
Qy 681 CAGTGTTCCTCATTTCT 740
Db 215 CAGTGTTCCTCATTTCT 156
Qy 741 TGCTAGCTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACTGCGG 800
Db 155 TGCTAGCTCTGCTCGCAATGAGGAGCATGTACCCAGGTACATCCATGAACTGCGG 96
Qy 801 CAGCAGTTGACTATATGCTTTCAGCTTTAAGGTTGTTGTTT 848
Db 95 GCAGCAGTTGACTTA-TGCTGTTTTCAGCTTTAAGGTTGTTGTTT 49

RESULT 13
AD122423
ID AD122423 standard; DNA; 1022 BP.
XX
XX AD122423;
XX
XX 22-APR-2004 (first entry)
XX
XX Rat liver differentially expressed cDNA seq id 233.
XX
XX hepatotropic; antiinflammatory; virucide; cytostatic; gene therapy;
XX differentially expressed; liver; toxin; liver disorder;
XX biliary cirrhosis; X-linked adrenoleukodystrophy; Zellweger syndrome;
XX hepatorenal syndrome; hepatitis; hepatocarcinoma; metabolic response;
XX toxicological response; ss; EST; expressed sequence tag; rat.
XX
XX Rattus sp.
XX
XX US2003165854-A1.
XX
XX 04-SEP-2003.
XX
XX 05-DEC-2001; 2001US-00006285.
XX
XX 05-DEC-2000; 2000US-0251986P.
XX
XX (CUNN/) CUNNINGHAM M J.
XX (KASE/) KASER M R.
XX
XX Cunnigham MJ, Kaser MR;
XX WPI; 2003-863697/80.
XX
XX New combination comprising a number of cDNAs that are differentially
XX expressed in a liver treated with a toxin, useful for diagnosing, staging
XX or treating liver disorders (e.g. biliary cirrhosis, hepatitis or
XX hepatocarcinoma).
XX
XX Claim 1; SEQ ID NO 233; 28pp; English.
XX
XX The invention describes a combination comprising a number of cDNAs that
XX are differentially expressed in a liver treated with a toxin and are
XX selected from any of the 514 cDNAs listed in the specification, or their
XX complements. The combination is useful in diagnosing, staging or treating
XX a liver disorder (e.g. biliary cirrhosis, X-linked adrenoleukodystrophy,
XX Zellweger syndrome, hepatorenal syndrome, hepatitis or hepatocarcinoma),
XX in monitoring diagnostic and therapeutic applications, in detecting
XX metabolic and toxicological responses, and in elucidating drug mechanism
XX of action. This sequence represents a cDNA differentially expressed in
XX liver tissues in response to treatment with a toxin.
XX
XX Sequence 1022 BP; 309 A; 160 C; 229 G; 250 T; 0 U; 74 Other;

Query Match 57.1%; Score 498.4; DB 10; Length 1022;
Best Local Similarity 87.3%; Pred. No. 6.4e-128;
Matches 580; Conservative 0; Mismatches 81; Indels 3; Gaps 3;
Qy 52 AAAGCTTGCCGAACATAAGCAAGAATGCTTCTGCTGCTGTTGGAGACCAAGGAATAAA 111
Db 104 AAAGCTTGCCGAACATAAGCAAGAATGCTTCTGCTGCTGTTGGAGACCAAGGAATAAA 163
Qy 112 GCAAGATCTTATCCACAGACT-CCAGGCCATATCTTGAAGCAATCTGTAAGGAGGCA 170
Db 164 ACAAGATCTTATCAATAGGCTACAGGCCATATCTTGAAGCAATCTGTAAGGAGGCA 223
Qy 171 ATGAAGAGATCTACTTGGGAGATGAAACAGAGGAGAAAGAAACAAAGCCATTGAGCTCC 230
Db 224 ATGAAGAGATCTACTTGGGAGATGAAACAGAGGAGAAAGAAACAAAGCCATTGAGCTCC 283
Qy 231 CTGTCAAAGAGAAAGAACCCCTCTGAAAACTGTGTGATGTGGCAGCAGAGAGAAAGATGG 290
Db 284 CTGTCAAAGAGAAAGAACCCCTCTGAAAAAGTTGTTGATATGGCATCAGAAAAAGAGGTG 343
Qy 291 TGAAAAATTACATCTGAAATACACAGACTGAGAGATGCAGAGAGGGCTGACGATTCA 350
Db 344 TAAAAATTACATCTGGAATACCTCAAACTGAGAGATGCAGAGAGGGCTGACGATTCA 403
Qy 351 ATGTACCTGTGAGCTTGGAGAGTAAAGAAAGCTGCTCG-GGCAGCTAGGTTTGGGATTCT 409
Db 404 ATGTGCTGTAGCTTGGAGAGTAAAGAGCTGCTCGCGCAGCGAGGTTTGGAAATTTCT 463
Qy 410 TCAGTTCCAAACAAAGGTTTATCATCTGACACCAAGCCAATGGTTAACCTGGATAAACTA 469
Db 464 TCAGTTCCAAACAAAGGTTTATCATCTGACACCAAGCCAATGGTTAACCTGGATAAACTA 523
Qy 470 AAGGAAGAGCTCAAGAGTTTGGTTGATGTCTTCAATCTCCAGAAAGTCTGAAGAT 529
Db 524 AAGGAAGAGCAGAGATTTGGTTTGAATGTCTTCCATCTCTAAGAAAGTCTGAGGAT 583
Qy 530 GATGAGAACTGAAAAAGAGAGGAGCGATTGGGATTGTCAAGTTTCAGCTGGAACCT 589
Db 584 GATGAGAGCTGAGAAACGGAAGAGAGATTGGATTGTGCAAGTTTCAGCTGGAACCT 643
Qy 590 GGAACCAACAGAGATACAGAGGCAAGAGAAAGAGAGAGCGCTTTGGGATTGCC 649
Db 644 GGAACCAACAGAGATACAGAGGCAAGAGAAAGAGAGAGCGCTTTGGGATTGCC 703
Qy 650 TGATGAAAGTTCCCTG-ATACCTTCTGTTCTCCAGTGTTCATTTCTCTCTCTCTCT 708
Db 704 TAATGAAAGTTCCCTGCTTCTGCCCCACCCCATAGTGGTTTCCATTTCTCAGATTCTT 763
Qy 709 TGGT 712
Db 764 GGT 767
RESULT 14
AAC00738
ID AAC00738 standard; cDNA; 471 BP.
XX
XX AAC00738;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 736.
XX
XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-00200610.


```
PR 09-JUN-2003; 2003US-0476641P.
PR 08-JUL-2003; 2003US-0485218P.
PR 08-JUL-2003; 2003US-0485223P.
PR 08-JUL-2003; 2003US-0485224P.
PR 08-JUL-2003; 2003US-0485325P.
PR 14-JUL-2003; 2003US-0486446P.
PR 14-JUL-2003; 2003US-0486480P.
PR 15-JUL-2003; 2003US-0486891P.
PR 15-JUL-2003; 2003US-0486960P.
PR 08-AUG-2003; 2003US-0493341P.
PR 08-AUG-2003; 2003US-0493370P.
PR 08-AUG-2003; 2003US-0493573P.
PR 08-AUG-2003; 2003US-0493577P.
XX
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX
XX Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
PI Halenbeck RF, Huang MW, Kothakota S, Haishan L, Linnemann T;
PI Pierce K, Wang Y, Wong JGP, Wu G, Zhang H;
XX
XX WPI; 2004-348438/32.
XX
XX New nucleic acid molecule for diagnosing, preventing or treating diseases
PT such as proliferative (e.g. cancer), inflammatory, immune, metabolic,
PT genetic, bacterial and viral diseases.
XX
XX Claim 1; SEQ ID NO 789; 428pp; English.
XX
XX The present invention relates to an isolated nucleic acid molecule
CC encoding a polypeptide which is believed to be cytostatic,
CC antiinflammatory, immunosuppressive, antibacterial and virucidal. The
CC composition and methods are useful for diagnosing, preventing and
CC treating diseases such as proliferative (e.g. cancer), inflammatory,
CC immune, metabolic, genetic, bacterial and viral diseases. The present
CC sequence represents a human secreted protein encoding sequence. The
CC present sequence is available on WIPWEB and is not in the specification.
XX
XX Sequence 558 BP; 202 A; 92 C; 147 G; 117 T; 0 U; 0 Other;
```

```
Query Match 51.7%; Score 451.6; DB 12; Length 558;
Best Local Similarity 93.2%; Pred. No. 5.1e-115;
Matches 533; Conservative 0; Mismatches 19; Indels 20; Gaps 5;
```

```
QY 82 TGCTCGTGGTTTGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGACTCCAGGCATA 141
DB 6 TTCTTGTGGTTTGAGACCAAGGGAATAAAGCAAGATCTTATCCACAGACTCCAGGCATA 65
QY 142 TCTTGAAGAACATCTCAAGAGGAGGCAAAATGAAGAGATGTACTGGGAGATGAACACAGA 201
DB 66 TCTTGAAGAACATCTCAAGAGGAGGCAAAATGAAGAGATGTACTGGGAGATGAACACAGA 122
QY 202 GGAAGAGAAACAAAGCCCATTTAGCTCCCTGTCAAAAGAGAAAGAACCCCTG-AAAAAA 260
DB 123 AGAAGAAAAA-----AACCCCTTGTCAAAATAGGAAGAACCCCTGAAAAAAA 169
QY 261 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATACATCTGAAATACCACAGACTG 320
DB 170 CTGTTGATGTGGCAGCAGAGAGAAAGTGGTGAATAATACATCTGAAATACCACAGACTG 229
QY 321 AGAGAATGCAGAGAGAGGCTGAACGATTCAATGTACCTGTGAGCTTGGAGAGTAAGAAAG 380
DB 230 AGAGAATGCAGAGAGAGGCGCAAAATCAATGTACCTGTGAGCTTGGAGAGTAAGAAAG 289
QY 381 CTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGTTCTCATCTGATA 440
DB 290 CTGCTCAGGCAGCTAGGTTTGGGATTTCTTCAGTTCCAAACAAAGGTTCTCATCTGATA 346
QY 441 ACAACCTATGTTAACTTGGATTAAGCTGAAGGAAGAGCTCAAGATTTGGTTGAAATG 500
DB 347 ACACACCTATGGTTAACTTGGATTAAGCGGAAGGAAGAGCTCAAGATTTGGTTGAAATG 406
QY 501 TCTCTTCAATCTCCAGAAAGTCTCAAGATGATGAAGAACTGAAAGAGAGGAGCGCAT 560
DB 501 TCTCTTCAATCTCCAGAAAGTCTCAAGATGATGAAGAACTGAAAGAGAGGAGCGCAT 560
```

```
Search completed: October 4, 2005, 19:50:03
Job time : 563.261 secs
```

```
Db 407 TCTCTTCAATTTCCAGAAAGTCTGAAGATGATAAGAAAACTGAAAAAGAGGAGCGAT 466
QY 561 TTGGGATTTGTCACAAAGTTTCAGCTGGAACTGGAAACCAACAGAGGATACAGAGGCAAGAGA 620
DB 467 TTGGGATTTGTCACAAAGTTTCAGCTGGAACTGGAAACCAACAGAGGATACAGAGGCAAGAGA 526
QY 621 GGAAGAGAGCAGAGCGCTTTGGGATTTGCCTGA 652
DB 527 GGAAGAGAGCAGAGCGCTTTGGCATTGCCTGA 558
```